FACTOR ANALYSIS OF DATA MATRICES

PART V

Paul Horst

Project Staff

Fern Thomas
Helen Ranck

This Study Was Supported in Part by

Office of Naval Research Contract Nonr -477(33)

and

Public Health Research Grant MH00743-08

Principal Investigator: Paul Horst

March 1964

University of Washington
Seattle, Washington

FACTOR ANALYSIS OF DATA MATRICES

FOREWORD

This is Part V of a series of reports on rationales and techniques of matrix factoring which play an important role in multivariate analysis techniques. Indeed, it may well be said that all adequate models and methods of multivariate analysis are special cases of matrix factoring techniques. The more traditional methods of factor analysis, in particular, are special cases of more general matrix factoring techniques, as are also all multiple regression models.

TABLE OF CONTENTS

PART V

CHAPTER 17

PRIMARY FACTOR MATRICES FROM HYPOTHESES

We have seen in the previous chapters that we may calculate, in a wide variety of ways, a factor loading matrix whose major product moment gives a lower rank approximation to the correlation or covariance matrix obtained from data matrices. We also know from Chapter 4 that, if any of these matrices obtained by a particular computational method were postmultiplied by a square orthonormal matrix, the major product moment of the resulting matrix would be the same as the major product moment of the matrix prior to multiplication by the orthonormal matrix. This, of course, is because in the major product moment the square orthonormal matrix is multiplied by its transpose to yield the identity matrix.

We also saw in Chapter 4 that, if we regard the factor loading matrix as one of the factors in the product of the two matrices which purports to approximate the data matrix, the same situation prevails. Suppose we have some approximation to a factor score matrix, postmultiplied by the transpose of a factor loading matrix, as a lower rank approximation to the data matrix. We may then also have the factor score matrix postmultiplied by an orthonormal matrix to yield another factor score matrix, and the transpose of the factor loading matrix premultiplied by the transpose of the same orthonormal matrix. Then the product of these two transformed matrices would be exactly the same as that of the original matrices.

Furthermore, we learned that these transformation matrices need not be square orthonormal, but that we may have a more general situation. We may have the factor score matrix postmultiplied by some square basic matrix, and the transpose of the factor loading matrix premultiplied by the inverse of this same matrix. Then the major product of the two transformed matrices would be the same as the major product of the two original matrices. This is because the square basic matrix multiplied by its inverse yields the identity matrix.

It is clear then, that we may have a multiply infinite number of factor loading matrices whose major product moments give identical results. Similarly, we may have an infinite number of pairs of factor score and factor loading matrices, whose major products give identical results. The question then arises as to which of these pairs of factor loading and factor score matrices is best, in some defined sense.

In this chapter we shall consider only factor loading matrices. We shall attempt to achieve some transformation of the arbitrary matrix so that the new factor loading matrix will have the following characteristics. First, for each factor loading vector, only a relatively small number of the variables shall have high loadings, and the remainder should have small loadings. Second, each variable shall have loadings in only a few of the factors. Third, for any given pair of factors, a number of the variables shall have small loadings in both factors. Fourth, for any given pair of factors, some of the variables shall have high loadings in one factor and low in the other, while other variables shall have high loadings in the second factor but not in the first. Fifth, for any given pair of factors, very few of the variables shall have high loadings in both.

These are the conditions which Thurstone (1947) has formulated as the "simple structure" criteria. We shall therefore refer to transformations which attempt to achieve these objectives as simple structure transformations, and we shall refer to the transformed factor loading matrix as a simple structure factor loading matrix.

There are, in general, three kinds of methods which have been used for transforming arbitrary factor matrices to simple structure matrices. The first of these is by means of graphical methods of rotation. In this method, every vector of factor loadings is plotted against every other vector, and by inspection of the plots, rotations or transformations are made two at a time. This is the oldest of the methods. It was developed by Thurstone (1947) and has been used extensively. The chief disadvantages of the method are that it is extremely time-consuming; it is not adapted to objective computational routines; and a great deal of personal judgment is left to the individual who does the plotting and the transformations.

The second procedure is based on *a priori* hypotheses. Here the investigator has some *a priori* hypothesis as to which variables should have high loadings in which factors, and which should not. This we may call the hypothesis method of transformation.

The third type of method involves analytical or mathematical criteria for transformation. These methods adopt certain mathematical functions of the transformed factor loadings which are to be optimized. They are called the analytical methods of transformation.

While the graphical methods have been extensively used in the past, they have been falling more and more into disuse as more objective methods have been developed. Therefore, these methods will not be discussed in this book. In this chapter, we shall give consideration to those methods which start with some hypothesis as to which factor loading should be sizable in which variables. In the next chapter, we shall consider some analytical methods which do not depend on *a priori* hypotheses.

In any case, it must be emphasized, as Thurstone (1947) has done so frequently, that the extent to which simple structure can be achieved by any of the methods is definitely limited by the nature of the data to be analyzed, and that simple structure must be inherent in the data if any of the methods is to reveal it in the transformed factor matrices.

17.1 Characteristics of the Hypothesis Methods

The hypothesis methods of transformation are similar in several respects. First, they begin with some arbitrary factor loading matrix--that is, with some factor loading matrix computed by one of the various methods outlined in the previous chapters, or by other methods. Second, they are based on some hypothesis as to which of the tests have high loadings in which factors. The methods differ essentially in the types of transformation matrices employed.

17.1.1 The Arbitrary Factor Matrices. The methods all begin with some arbitrary factor matrix such as the centroid, the multiple group, the group centroid, or the principal axis solution. In the methods we have discussed, it will be recalled that they are all special cases of the rank reduction method. It is, however, not necessary that the arbitrary factor matrices be special cases of the rank reduction method. They may be based on some clustering or B-coefficient methods, such as described by Holzinger and Harman (1941).

The computational routines of the hypothesis methods differ essentially according to which particular type of arbitrary matrix is used. It should be emphasized at this point that most of the procedures we have outlined have had as their primary objective the finding of a factor loading matrix which, with a minimum number of factors, will give the best approximation to the correlation or covariance matrix. A major concern has been to find the smallest number of factors which, with a satisfactory degree of accuracy, can reproduce the correlation or covariance matrix and, indirectly, the data matrix. This objective recognizes the finding of a lower rank best approximation to a data matrix as fundamentel to all scientific investigations.

17.1.2 The Hypothesis Matrix. The group of methods considered in this chapter depend on having some a priori hypotheses as to which tests should have high loadings in which factors, and which tests should have low loadings. These methods are characterized by the specification of a binary hypothesis matrix. This binary matrix has a 1 in the *ij*th position if the *i*th test has a high loading in the *j*th factor. Otherwise, it has a 0 in this position.

This binary hypothesis matrix can be made up after the variables are assembled. However, Thurstone (1947), Guttman (1952), and other have emphasized that it is better first to make up an hypothesis matrix and then to attempt to specify variables which will satisfy the hypothesis.

The ideal binary hypothesis matrix would be one in which there is only a single 1 in each row, and roughly an equal number of 1's in each column. This would be a nonoverlapping hypothesis matrix. However, there is nothing to prevent one from having a more complex hypothesis so that he may have several 1's in each row for some of the variables, indicating that he believes the variables have loadings on more than one of the factors.

It will be recalled that a binary matrix is also involved in the multiple group method of factor analysis, and that binary vectors are involved in the group centroid method. In those methods, however, it was considered desirable, but not mandatory, that these binary matrices or vectors represent plausible hypotheses as to the factor loadings to be found.

17.1.3 The Transformation Matrix. The methods described in this chapter all require a decision as to what particular conditions are to be satisfied by the transformation matrix. Usually it is considered desirable to have a transformation matrix which is normal by columns. That is, the minor product moment of the transformation matrix should have unity in the diagonals. This restriction is required so that the factor vectors of the new transformed factor matrix will be comparable to one another. More technically, such a transformation provides the basis for finding the correlation among the new reference axes from which the new factor loadings are measured.

This concept, however, involves us with geometric and trigonometric symbolic systems which we wish to avoid in this book. We have attempted to restrict ourselves to algebraic and arithmetic concepts. This has been in the belief that the traditional random mixture of various types of symbolic mathematical systems does not yield a better understanding of the phenomena under study, unless one is already thoroughly familiar with the standard symbolic systems and the interrelationships among them. The assumption in this book is that many readers are not thoroughly familiar with these overlapping and interrelated symbolic systems.

The methods differ with respect to the type of transformation involved.

These may be square orthonormal transformations. If the transformations are not orthonormal, they are called oblique. Therefore, a choice between orthonormal and oblique transformations is available. Although the term "oblique" is not particularly appropriate, we shall continue to use it since it is well-established in the literature.

The methods differ also in restrictions which may be placed on the transformed simple structure factor loading matrix. We shall describe one method. More specifically, one type of transformation is such that, for all variables which have 0's in a given column of the hypothesis binary matrix, the sum of their loadings in the simple structure matrix for the corresponding factor vector is 0.

17.2 Kinds of Methods

In this chapter we shall outline three methods which differ with respect to the type of arbitrary matrix on which they are based. A fourth method will yield a zero sum for assumed zero loadings in each factor. A fifth method uses a square orthonormal transformation.

17.2.1 The Multiple Group Matrix. This method begins with an arbitrary factor matrix obtained by the multiple group method. We shall assume that the binary matrix used in the multiple group method was actually an hypothesis matrix, and that it is the one which the transformed simple structure matrix is to resemble as closely as possible. This particular type of arbitrary factor loading matrix is regarded as a special case because certain computational simplifications are possible when the binary grouping matrix and the binary hypothesis matrix are the same.

17.2.2 The Principal Axis Arbitrary Matrix. The second method we shall outline uses the principal axis or basic structure factor loading matrix as the arbitrary matrix. In many cases when high speed computers have been used, the solution will be a principal axis or basic structure factor matrix. The computations for the transformed simple structure matrix are simplified because the principal axis solution yields an orthogonal factor matrix.

17.2.3 The Arbitrary Factor Matrix. This method may be regarded as a generalization of other methods, of which the multiple group and the principal axis arbitrary matrices are special cases. However, the solution does not depend on any peculiar properties of the factor loading matrix, such as orthogonality, or on the identity of the hypothesis and the grouping binary matrix, as in the multiple group method. The method does not depend on how the arbitrary factor matrix was determined, and requires only the construction of the binary hypothesis matrix.

17.2.4 The Zero Partial Sum Method. This method is independent of the particular type of arbitrary matrix on which it is based. We may begin with any of the solutions discussed in the previous chapters. In any case, the transformation of the arbitrary matrix yields a simple structure matrix such that those variables which have zeros in an hypothesis vector will yield a zero sum for the corresponding factor loadings. This does not mean necessarily that the simple structure criteria may be well satisfied, or that the binary hypothesis matrix may be reasonably well approximated by the transformed matrix. For example, it is quite possible that even though partial sums are 0, the elements going into the sum may still vary greatly. Furthermore, appreciable or high loadings may not appear in the simple structure matrix to correspond with 1's in the hypothesis matrix. The results yielded by this or any simple structure solution are as much a function of the data themselves as of the particular method used.

17.2.5 The Orthonormal Transformation. In the methods discussed above, the type of transformation--whether orthonormal or oblique--was not mentioned. It will be recalled that only if the transformation of a factor loading matrix is orthonormal will the major product moment of the original and the transformed factor loading matrix be the same. There is nothing in the computational procedures of the methods just discussed to guarantee that the transformation matrix will be orthonormal. In general, it will be oblique. It should be emphasized here that this is not a serious objection because, as we have indicated in Chapter 4, the main objective is not to reproduce the correlation or covariance matrix with the major product moment of a factor loading matrix, but rather to reproduce the original or rescaled data matrix as closely as possible by the major product of a factor score and a factor loading matrix. Therefore, if we get a given factor loading matrix and transform it with some nonorthonormal or oblique matrix, we can always transform the factor score matrix corresponding to it by the inverse of this transformation, so that the major product of the two will be the same as the major product of the original factor score and factor loading matrices.

However, there has been considerable insistence among some research workers that transformations be orthonormal, and it is of interest to consider a method which will guarantee an orthonormal solution based on the binary hypothesis matrix. The chief advantage of the orthonormal transformation procedures is that in certain cases, as, for example, the principal axis method, one can be sure that the factor scores are uncorrelated, or what amounts to the same thing, that the factor score matrix is an orthogonal matrix.

17.3 The Multiple Group Factor Matrix

17.3.1 Characteristics of the Method. This method assumes that we have adopted the multiple group method of factoring the correlation matrix. As will be recalled, this method begins with the multiplication of the correlation covariance matrix by a grouping binary matrix. In the technique of this section, the grouping binary matrix is the same as the binary hypothesis matrix.

However, the procedure is such that one does not need to carry out all of the computations for the multiple group factor matrix. Because the grouping and the hypothesis binary matrices are the same, one can omit some of the computations for the multiple group matrix. Distinguishing characteristics of this method are that one begins with only a partial solution of the multiple group matrix, and that the computations are somewhat simpler than for other hypothesis methods discussed in this chapter.

17.3.2 Computational Equations

17.3.2a Definition of Notation

$\underline{R}$ is the correlation matrix.

$\underline{f}$ is the binary hypothesis matrix.

$\underline{H}$ is the transformation matrix.

$\underline{r}$ is the correlation of the reference axes.

$\underline{b}$ is the simple structure factor matrix.

17.3.2b The Equations

$$F = R\,f \tag{17.3.1}$$

$$S = F'\,f \tag{17.3.2}$$

$$t\,t' = S \tag{17.3.3}$$

$$G = F'\,F \tag{17.3.4}$$

$$C = G^{-1}\,S \tag{17.3.5}$$

$$g = S\,C \tag{17.3.6}$$

$$\gamma = g\,C \tag{17.3.7}$$

$$D = D_{\gamma}^{-\frac{1}{2}} \tag{17.3.8}$$

$$H = t'\,C\,D \tag{17.3.9}$$

$$r = D\,\gamma\,D \tag{17.3.10}$$

$$b = F\,(C\,D) \tag{17.3.11}$$

17.3.3 Computational Instructions. We begin the computations with the correlation or covariance matrix, rather than with the multiple group factor matrix, because we shall omit the final steps of the multiple group solution. We assume that the binary matrix $\underline{f}$ has been constructed. It has 1's for those tests in each factor which are assumed to have high loadings, and 0's for all of the others.

The first step, as in the multiple group method, is indicated in Eq.

(17.3.1). Here we postmultiply a correlation matrix by the hypothesis binary matrix f. Obviously, this merely serves to sum those columns of the correlation matrix corresponding to the 1's in the vectors of the f matrix on the right of Eq. (17.3.1).

The next step, as in the multiple group method, is to premultiply the f matrix by the transpose of the F matrix calculated in Eq. (17.3.1). This is indicated in Eq. (17.3.2) by the S matrix on the left of the equation. This, in effect, adds rows of the F matrix from Eq. (17.3.1), corresponding to the 1's in the vectors of the f matrix, to give a symmetric matrix, S, in Eq. (17.3.2).

Eq. (17.3.3) indicates a triangular factoring of the matrix S of Eq. (17.3.2). This triangular factoring is not necessary unless one actually wishes to see the transformation matrix which is indicated in Eq. (17.3.9). For most of the parameters of interest in a factor analysis, this transformation matrix, as such, is not required. It is only to make the analysis complete that it might be included.

Eq. (17.3.4) is the minor product moment of the F matrix calculated in Eq. (17.3.1). This is indicated on the left by G.

Eq. (17.3.5) requires the calculation of the inverse of the G matrix calculated in Eq. (17.3.4). The matrix designated as C in Eq. (17.3.5) is the S matrix of Eq. (17.3.2) premultiplied by the inverse of the G matrix given in Eq. (17.3.4).

The next step is to calculate the product indicated in Eq. (17.3.6). Here we premultiply the C matrix of Eq. (17.3.5) by the S matrix of Eq. (17.3.2). This we designate as g.

We next require the $\underline{\gamma}$ matrix indicated in Eq. (17.3.7). This is obtained by premultiplying the $\underline{C}$ matrix of Eq. (17.3.5) by the $\underline{g}$ matrix of Eq. (17.3.6).

The computation of a diagonal matrix is indicated by Eq. (17.3.8). This is simply a matrix whose elements are the reciprocal square roots of the diagonal elements of the $\underline{\gamma}$ matrix given by Eq. (17.3.7).

The computation of the $\underline{H}$ matrix, which is the transformation matrix for the multiple group factor matrix, is indicated in Eq. (17.3.9). This is the triple product, from left to right, of the upper triangular factor of the $\underline{S}$ matrix in Eq. (17.3.3) by the $\underline{C}$ matrix of Eq. (17.3.5) by the diagonal matrix of Eq. (17.3.8). As indicated above, this matrix is not used in the computation of the simple structure factor loading matrix. It is used in Chapter 21 for the calculation of general factor parameters, but if these are not desired it need not be calculated.

The calculation of a correlation matrix is indicated in Eq. (17.3.10). This matrix is the minor product moment of the $\underline{H}$ matrix of Eq. (17.3.9). It could be calculated as such from Eq. (17.3.9), but it is simpler to calculate it from Eqs. (17.3.7) and (17.3.8). This matrix is of interest because it shows the extent to which the reference vectors of the transformation matrix are correlated. It is precisely a correlation matrix of the simple structure reference axes. In most factor analyses, both this matrix and its normalized inverse are calculated to indicate the extent to which the transformation departs from orthonormality. More particularly, these matrices are useful for purposes of further analysis, as indicated in Chapter 2. The matrix is of interest also because, while in general simple structure factor matrices on the same variables are supposed to be relatively invariant from one type of sample to another, the correlation matrix of the simple structure reference axes as given in Eq. (17.3.10) may be regarded as characterizing the particular sample on which the analysis is based. It may vary greatly from one sample to another.

The calculation of the simple structure factor loading matrix is given in Eq. (17.3.11). As indicated on the right of this equation, we compute first a matrix which is the product of the $\underline{C}$ matrix of Eq. (17.3.5) postmultiplied by the diagonal matrix of Eq. (17.3.8). This matrix is then postmultiplied into the $\underline{F}$ matrix of Eq. (17.3.1).

The characteristic of this $\underline{b}$ matrix in Eq. (17.3.11) is that it should represent as nearly as possible in the least square sense the simple structure hypothesis matrix $\underline{f}$ which has been scaled on the left by a diagonal such that the transformation matrix $\underline{H}$ in Eq. (17.3.9) is normal by columns.

17.3.4 Numerical Example. We begin this numerical example with the correlation matrix of previous chapters.

We shall solve for only three factors; therefore the binary matrix consists of only three column vectors. The unit elements in these three vectors are taken, respectively, as the first three, the second three, and the third three.

Table 17.3.1 gives the correlation matrix postmultiplied by the binary matrix $\underline{f}$, as indicated by Eq. (17.3.1).

Table 17.3.2 gives the minor product $\underline{S}$ of the $\underline{f}$ and the $\underline{F}$ matrices, as indicated by Eq. (17.3.2).

Table 17.3.3 gives the minor product moment $\underline{G}$ of the $\underline{F}$ matrix, as indicated by Eq. (17.3.4).

Table 17.3.4 gives the inverse of the $\underline{G}$ matrix shown in Table 17.3.3.

Table 17.3.5 gives the product of the $\underline{S}$ matrix of Table 17.3.2 premultiplied by the $\underline{G}^{-1}$ matrix of Table 17.3.4. This product is the $\underline{C}$ matrix of Eq. (17.3.5).

In the body of Table 17.3.6 is given the product of the $\underline{C}$ matrix of Table 17.3.5, premultiplied by the $\underline{S}$ matrix of Table 17.3.2. This product is the $\underline{g}$ matrix of Eq. (17.3.6). The last line of Table 17.3.6 is obtained by calculating only the diagonal elements of the product $\underline{\gamma} = \underline{g}\,\underline{C}$ given by Eq. (17.3.7), and taking the reciprocal square roots of these elements as indicated by Eq. (17.3.8).

Table 17.3.7 is the simple structure factor loading matrix $\underline{b}$. This is the triple product of the matrix $\underline{F}$ given by Table 17.3.1, the matrix $\underline{C}$ given by Table 17.3.5, and a diagonal matrix $\underline{D}$ constituted from the elements of the last row of Table 17.3.6. This product is indicated by Eq. (17.3.11).

The computations indicated by Eqs. (17.3.3), (17.3.9), and (17.3.10) are not given in this numerical example.

Table 17.3.1 - Matrix $\underline{R}\,\underline{f} = \underline{F}$

2.59700	0.24900	0.95800
2.60400	0.30100	1.03900
2.54300	0.71500	0.95200
0.49500	2.26200	0.80100
0.29900	2.34500	0.48300
0.47100	2.33500	0.58400
0.91700	0.57700	2.18100
0.92700	0.37700	2.19500
1.10500	0.91400	2.06800

Table 17.3.2 - Matrix $\underline{F}'\,\underline{f} = \underline{S}$

7.74400	1.26500	2.94900
1.26500	6.94200	1.86800
2.94900	1.86800	6.44400

Table 17.3.3 - Matrix $\underline{F}'\,\underline{F} = \underline{G}$

23.46958	8.05789	14.75028
8.05789	18.04217	9.51620
14.75028	9.51620	17.97094

Table 17.3.4 - Matrix $\underline{G}^{-1}$

0.08804	-0.00167	-0.07138
-0.00167	0.07694	-0.03937
-0.07138	-0.03937	0.13508

Table 17.3.5 - Matrix $\underline{G}^{-1}\,\underline{S} = \underline{C}$

0.46919	-0.03358	-0.20345
-0.03173	0.45844	-0.11490
-0.20421	-0.11126	0.58641

Table 17.3.6 - Matrix $\underline{S}\,\underline{C} = \underline{g}$ and Vector of $\underline{D}\,1 = \underline{D}_{g\,C}^{-2}\,1$

2.99109	-0.00818	0.00847
-0.00818	2.93220	0.04042
0.00847	0.04042	2.96422

0.845	0.864	0.760

Table 17.3.7 - Simple Structure Factor Loading Matrix

$\underline{b} = \underline{F}\ (\underline{C}\ \underline{D})$

0.857	-0.069	0.004
0.845	-0.056	0.034
0.824	0.118	-0.031
-0.003	0.804	0.083
-0.028	0.874	-0.036
0.023	0.855	-0.016
-0.028	-0.008	0.780
-0.021	-0.089	0.802
0.057	0.131	0.671

17.4 The Principal Axis Factor Matrix

17.4.1 Characteristics of the Method. This method, like the previous one, is a special case of an arbitrary matrix resulting in simplified computations. Here the simplification results because the principal axis factor matrix is orthogonal by columns.

One of the most important characteristics of the method is that the computations do not call for the calculation of the inverse of any matrix except a diagonal matrix. This is far easier to calculate than the inverses of symmetric or other square matrices. The method is of considerable practical importance because, with the increasing availability of high speed computers, most of the arbitrary factor loading matrices calculated will be of the principal axis or basic structure type.

17.4.2 The Computational Equations

17.4.2a Definition of Notation

$\underline{a}$ is the principal axis factor matrix.

$\underline{\delta}$ is that part of the basic diagonal matrix of $\underline{R}$ corresponding to $\underline{a}$.

$\underline{r}$ is the correlation matrix of the reference axes.

$\underline{b}$ is the simple structure factor loading matrix.

17.4.2b The Equations

$$G = a' f \qquad (17.4.1)$$

$$C = \delta^{-1} G \qquad (17.4.2)$$

$$S = C' C \qquad (17.4.3)$$

$$D = D_{C'}^{-\frac{1}{2}} C \tag{17.4.4}$$

$$H = C\,D \tag{17.4.5}$$

$$r = D\,S\,D \tag{17.4.6}$$

$$b = a\,H \tag{17.4.7}$$

17.4.3 Computational Instructions. The computations in this method are relatively simple.

We begin with Eq. (17.4.1). This is simply the product of the transposed principal axis factor loading matrix postmultiplied by the binary hypothesis matrix $\underline{f}$.

Next we calculate the $\underline{C}$ matrix in Eq. (17.4.2). This is the $\underline{G}$ matrix of Eq. (17.4.1) premultiplied by the reciprocal of the basic diagonal corresponding to the first three principal axis factors. In other words, this basic diagonal includes only the basic diagonal elements of the correlation matrix corresponding to the three factors which have been solved for.

Next we calculate Eq. (17.4.3). This is the minor product moment of the $\underline{C}$ matrix calculated in Eq. (17.4.2).

We then calculate a diagonal matrix whose elements are the reciprocal square roots of the diagonal elements of $\underline{S}$ calculated in Eq. (17.4.3). This is indicated in Eq. (17.4.4).

Next we calculate the transformation matrix $\underline{H}$ as in Eq. (17.4.5). This is the matrix $\underline{C}$ of Eq. (17.4.2) postmultiplied by the diagonal matrix of Eq. (17.4.4). It will be noted that this computation was optional in the previous method. Here, however, it is required for further calculations, as will be indicated for Eq. (17.4.7).

The next step is to calculate the correlation of the simple structure reference axes, as indicated in Eq. (17.4.6). Here the matrix $\underline{S}$ of Eq. (17.4.3) is pre- and postmultiplied by the diagonal matrix of Eq. (17.4.4). As seen from Eqs. (17.4.3), (17.4.4), and (17.4.5), $\underline{r}$ could also have been calculated by taking the minor product moment of the $\underline{H}$ matrix in Eq. (17.4.5). However, this would have meant the multiplication of square matrices, whereas the computations in Eq. (17.4.6) require only operations on the symmetric matrix $\underline{S}$ by diagonal matrices.

Eq. (17.4.7) gives, finally, the simple structure factor loading matrix. This is the principal axis factor loading matrix postmultiplied by the transformation matrix $\underline{H}$ of Eq. (17.4.5). This matrix now is the best least square approximation to the scaled binary hypothesis matrix $\underline{F}$, in which the scaling is such as to make the column vectors of $\underline{H}$ in Eq. (17.4.5) normal.

17.4.4 Numerical Example. We take as numerical data for this illustration the principal axis factor loading matrix calculated in Chapter 8. This is repeated for convenience in Table 17.4.1, together with the first three basic diagonals in the top row. The binary hypothesis matrix is the same as given in the previous section. This hypothesis assumes that no variable has factor loadings in more than one factor.

Table 17.4.2 gives the product of the transpose of the matrix in Table 17.4.1 postmultiplied by the binary hypothesis matrix.

Table 17.4.3 is the product $\underline{C}$ of the matrix of Table 17.4.2, premultiplied by the inverse of a diagonal matrix constituted from the basic diagonal elements at the top of Table 17.4.1. These calculations are indicated by Eq.

(17.4.2).

Table 17.4.4 is the minor product moment of the matrix $\underline{C}$ calculated in Table 17.4.3.

Table 17.4.5 is the matrix obtained by postmultiplying the matrix $\underline{C}$ of Table 17.4.3 by the inverse of a diagonal matrix whose elements are from the diagonal of the matrix in Table 17.4.4. This gives the transformation $\underline{H}$ indicated by Eq. (17.4.5).

Table 17.4.6 is the simple structure factor loading matrix. It is the product of the principal axis factor matrix of Table 17.4.1, postmultiplied by the matrix $\underline{H}$ of Table 17.4.5. This product is indicated by Eq. (17.4.7).

The correlation matrix $\underline{r}$ of Eq. (17.4.6) is not calculated in this numerical example.

Table 17.4.1 - First Three Basic Diagonals and Principal Axes Factor Vectors

3.749		
2.050		
1.331		

0.717	0.493	0.350
0.740	0.478	0.322
0.773	0.296	0.406
0.556	-0.649	0.068
0.463	-0.744	0.181
0.518	-0.694	0.188
0.640	0.080	-0.588
0.615	0.166	-0.621
0.715	-0.034	-0.369

Table 17.4.2 - Matrix $\underline{G} = \underline{a' f}$

2.23000	1.53700	1.97000
1.26700	-2.08700	0.21200
1.07800	0.43700	-1.57800

Table 17.4.3 - Matrix $\underline{C} = \underline{\delta^{-1} G}$

0.59483	0.40998	0.52547
0.61805	-1.01805	0.10341
0.80992	0.32832	-1.18557

Table 17.4.4 - Matrix $\underline{S} = \underline{C' C}$

1.39177	-0.11942	-0.58374
-0.11942	1.31230	-0.27910
-0.58374	-0.27910	1.69240

Table 17.4.5 - Transformation Matrix $\underline{H} = \underline{C\, D_{C'C}^{-\frac{1}{2}}}$

0.50420	0.35788	0.40392
0.52389	-0.88869	0.07949
0.68653	0.28661	-0.91133

Table 17.4.6 - Simple Structure Matrix $\underline{b} = \underline{a\, H}$

0.860	-0.081	0.010
0.845	-0.068	0.043
0.824	0.130	-0.034
-0.013	0.795	0.111
-0.032	0.879	-0.037
0.027	0.856	-0.017
-0.039	-0.011	0.801
-0.029	-0.105	0.828
0.089	0.180	0.622

17.5 The Arbitrary Factor Matrix

17.5.1 Characteristics of the Method. This method is more general than the two previous methods in that it applies to any factor matrix, including the principal axis, centroid, grouping, and multiple group matrices as special cases. However, the method includes a slight modification of the previous two methods. This modification is desirable in most transformation solutions, although frequently it makes very little difference, and it has been omitted from the two previous methods in order to simplify computations.

In this method, the arbitrary factor loading matrix is normalized by rows--that is, by variables--before transformation operations begin upon it. The reason for this is that all of the tests will then be given equal weight in the transformation solution. In most of the arbitrary type solutions which are designed primarily to find the minimum number of factors which can adequately account for a correlation or data matrix, all of the variables do not account for the same amount of variance in the factor matrix. In other words, what have been called the communalities of the variables, or the sums of squares of factor loadings for a given variable, will vary considerably from one variable to another. Therefore it is considered desirable to normalize the rows, or to make all of the test vectors of unit length. The unit length vector will be recognized by some as a geometric concept. We shall not, however, develop this concept further, since it would contribute to a confusion of symbolic systems. From the algebraic or arithmetic point of view we can simply state that we wish each of the tests to carry unit weight in the determination of the transformation. It may be recalled that in Chapter 15, for the communality type scaling method, the $\underline{\alpha}$ factor loading matrices satisfy this condition. That is, for any given number of factors, the sums of squares of rows of the $\underline{\alpha}$ matrix are all equal to 1.

17.5.2 The Computational Equations

17.5.2a Definition of Notation

$\underline{a}$ is an arbitrary factor matrix.

$\underline{f}$ is the binary hypothesis matrix.

$\underline{r}$ is the correlation of the simple structure reference axes.

$\underline{H}$ is the simple structure transformation matrix.

$\underline{b}$ is the simple structure factor matrix.

17.5.2b The Equations

$$D = D_{a\,a'}^{-\frac{1}{2}} \tag{17.5.1}$$

$$A = D\,a \tag{17.5.2}$$

$$G = A'\,f \tag{17.5.3}$$

$$C = A'\,A \tag{17.5.4}$$

$$M = C^{-1}\,G \tag{17.5.5}$$

$$g = M'\,M \tag{17.5.6}$$

$$D = D_{g}^{-\frac{1}{2}} \tag{17.5.7}$$

$$H = M D \tag{17.5.8}$$

$$r = H' H \tag{17.5.9}$$

$$b = A H \tag{17.5.10}$$

17.5.3 Computational Instructions. In this method we assume that all of the computations for an arbitrary factor matrix have been completed and that we start with this arbitrary matrix, however arrived at.

The first step in the calculations is indicated in Eq. (17.5.1). Here we calculate a diagonal matrix whose elements are the reciprocal square root of the diagonal elements of the major product moment of the arbitrary matrix $\underline{a}$. This means, of course, that we must calculate the sums of squares of row elements for the arbitrary factor matrix $\underline{a}$. These are what are called the communalities of the variables based on the particular factoring solution.

The second step is to calculate the normalized factor loading matrix $\underline{A}$, as indicated in Eq. (17.5.2). This is given by premultiplying the arbitrary factor matrix by the diagonal matrix in Eq. (17.5.1).

Eq. (17.5.3) is the minor product of the matrix in Eq. (17.5.2) and the hypothesis matrix $\underline{f}$.

We then calculate, as shown in Eq. (17.5.4), the minor product moment of the matrix $\underline{A}$ given by Eq. (17.5.2). This we call $\underline{C}$.

Next we calculate a matrix $\underline{M}$ which is the $\underline{G}$ matrix of Eq. (17.5.3) premultiplied by the inverse of the matrix $\underline{C}$ in Eq. (17.5.4).

We then calculate Eq. (17.5.6), which is the minor product moment of the matrix calculated in Eq. (17.5.5). This we call $\underline{g}$.

Next we calculate the diagonal matrix of Eq. (17.5.7). This is simply a matrix whose diagonal elements are the reciprocal square roots of the diagonal elements of the matrix $\underline{g}$ calculated in Eq. (17.5.6).

We now calculate the transformation matrix $\underline{H}$ indicated in Eq. (17.5.8). This is the matrix $\underline{M}$ of Eq. (17.5.5) postmultiplied by the diagonal matrix of Eq. (17.5.7). The method of computation indicates that the column vectors of the $\underline{H}$ matrix are normal.

Eq. (17.5.9) gives the correlation of the primary reference axes as the minor product moment of the transformation matrix $\underline{H}$ given in Eq. (17.5.8).

Finally, Eq. (17.5.10) gives the simple structure factor loading matrix. This is obtained by postmultiplying the matrix of Eq. (17.5.2) by the transformation matrix $\underline{H}$ given in Eq. (17.5.9).

17.5.4 Numerical Example

Table 17.5.1 gives the centroid factor loading matrix calculated in Chapter 6, Section 3. The same binary hypothesis matrix is used as in the previous two methods.

Table 17.5.2 gives the group centroid factor matrix normalized by rows, as computed by Eqs. (17.5.1) and (17.5.2).

Table 17.5.3 is the minor product $\underline{G}$ of the normalized factor matrix and the binary data matrix, as computed by Eq. (17.5.3).

Table 17.5.4 is the minor product moment $\underline{C}$ of the normalized factor matrix, as indicated by Eq. (17.5.4).

Table 17.5.5 is the inverse $\underline{C}^{-1}$ of the matrix of Table 17.5.4.

Table 17.5.6 is the matrix $\underline{M}$ calculated from the matrices of Tables 17.5.3 and 17.5.5, as shown in Eq. (17.5.5).

Table 17.5.7 is the transformation matrix H obtained by normalizing the columns of M in Table 17.5.6 by means of Eqs. (17.5.7) and (17.5.8).

Table 17.5.8 is the simple structure factor matrix b obtained from the matrices of Tables 17.5.2 and 17.5.7, indicated by Eq. (17.5.10). It is to be noted that the matrix b may be premultiplied by the inverse of the diagonal matrix of Eq. (17.5.1). This procedure is preferred by some factor analysts.

The correlation matrix r of Eq. (17.5.9) has not been calculated.

Table 17.5.1 - Group Centroid Factor Matrix $\underline{a}$

0.933	-0.068	0.002
0.936	-0.048	0.033
0.914	0.115	-0.035
0.178	0.840	0.073
0.107	0.885	-0.046
0.169	0.870	-0.027
0.330	0.165	0.777
0.333	0.087	0.800
0.397	0.283	0.667

Table 17.5.2 - Normalized Factor Matrix $\underline{A}$

0.99735	-0.07269	0.00214
0.99807	-0.05118	0.03519
0.99146	0.12475	-0.03797
0.20656	0.97476	0.08471
0.11987	0.99145	-0.05153
0.19060	0.98120	-0.03045
0.38366	0.19183	0.90333
0.38237	0.09990	0.91860
0.48052	0.34254	0.80732

Table 17.5.3 - Matrix $\underline{G} = \underline{A' f}$

2.98688	0.51703	1.24654
0.00087	2.94741	0.63426
-0.00064	0.00273	2.62925

Table 17.5.4 - Matrix $\underline{C} = \underline{A' A}$

3.59151	0.78369	1.09087
0.78369	3.08345	0.53650
1.09087	0.53650	2.32504

Table 17.5.5 - Matrix $\underline{C}^{-1}$

0.33511	-0.06023	-0.14333
-0.06023	0.34870	-0.05220
-0.14333	-0.05220	0.50939

Table 17.5.6 - Matrix $\underline{M}$ = $\underline{C^{-1} G}$

1.00098	-0.00466	0.00268
-0.17957	0.99649	0.00883
-0.42848	-0.22658	1.12755

Table 17.5.7 - Transformation Matrix $\underline{H}$ = $\underline{M\, D_{M'M}^{-\frac{1}{2}}}$

0.90706	-0.00456	0.00237
-0.16273	0.97510	0.00783
-0.38828	-0.22171	0.99997

Table 17.5.8 - Simple Structure Factor Matrix $\underline{b}$ = $\underline{A\, H}$

0.916	-0.076	0.004
0.900	-0.062	0.037
0.894	0.126	-0.035
-0.004	0.931	0.093
-0.033	0.978	-0.043
0.025	0.963	-0.022
-0.034	-0.015	0.906
-0.026	-0.108	0.920
0.067	0.153	0.811

17.6 The Zero Partial Sum Transformation

17.6.1 Characteristics of the Method. As indicated earlier in this chapter, the rationale of this procedure is to find a transformation such that, for any given factor vector, the factor loadings in the simple structure matrix corresponding to the zero elements in the corresponding vector of the binary hypothesis matrix shall add up to 0. This restriction can be imposed on any oblique solution for any arbitrary factor matrix. It applies equally well to the multiple group, principal axis, and the other methods.

In general, this additional restriction tends to give smaller values for the near-zero elements. However, there might still be considerable dispersion about the mean of 0 for these hypothesized zero elements. In any case, the computations, as will be seen, are somewhat more involved than they are in the methods previously considered.

17.6.2 The Computational Equations

17.6.2a Definition of Notation

$\underline{a}$ is an arbitrary factor matrix.

$\underline{f}$ is the binary hypothesis matrix.

$\underline{r}$ is the matrix of correlations of the simple structure reference axes.

$\underline{H}$ is the simple structure transformation matrix.

$\underline{b}$ is the simple structure factor matrix.

17.6.2b The Equations

$$S = a' a \qquad (17.6.1)$$

$$t\, t' = S \qquad (17.6.2)$$

$$\alpha = a\, t'^{-1} \qquad (17.6.3)$$

$$Y = \alpha'\, 1 \qquad (17.6.4)$$

$$W = \alpha'\, f \qquad (17.6.5)$$

$$U = Y\, 1' - W \qquad (17.6.6)$$

$$V = U\, D_{U'U}^{-\frac{1}{2}} \qquad (17.6.7)$$

$$Z = V\, D_{V'W} \qquad (17.6.8)$$

$$C = W - Z \qquad (17.6.9)$$

$$M = t'^{-1}\, C \qquad (17.6.10)$$

$$\gamma = M'\, M \qquad (17.6.11)$$

$$D = D_{\gamma}^{-\frac{1}{2}} \qquad (17.6.12)$$

$$H = M\, D_{\gamma}^{-\frac{1}{2}} \qquad (17.6.13)$$

$$r = D\, \gamma\, D \qquad (17.6.14)$$

$$b = a\, H \qquad (17.6.15)$$

17.6.3 Computational Instructions. In this method we may begin with any arbitrary factor loading matrix and operate directly upon it, or we may first normalize the rows of the factor loading matrix as we did in the method of Section 17.5. In either case, we have a binary hypothesis matrix $\underline{f}$, as in the previous three methods.

Eq. (17.6.1) gives the initial computations, which consist of the minor product moment of the factor loading matrix. If this happens to be a principal axis matrix, then, of course, the matrix $\underline{S}$ on the left will be a diagonal matrix.

We then indicate the triangular factoring of the $\underline{S}$ matrix of Eq. (17.6.1) by Eq. (17.6.2).

The next step is to postmultiply the factor loading matrix $\underline{a}$ by the inverse of the upper triangular factor of the $\underline{S}$ matrix in Eq. (17.6.2). This is indicated in Eq. (17.6.3).

Next we calculate a vector as indicated in Eq. (17.6.4). This is simply a vector whose elements are the sums of column elements of the $\underline{\alpha}$ matrix in Eq. (17.6.3). It is, of course, the transpose of the $\underline{\alpha}$ matrix of Eq. (17.6.3) postmultiplied by a unit vector.

Next we calculate the minor product of the $\underline{\alpha}$ matrix of Eq. (17.6.3) by the hypothesis binary matrix $\underline{f}$, as indicated in Eq. (17.6.5). This is the matrix $\underline{W}$ on the left of Eq. (17.6.5).

We then calculate the matrix $\underline{U}$ as indicated in Eq. (17.6.6). Each column of the $\underline{U}$ matrix in Eq. (17.6.6) is obtained by subtracting from the $\underline{Y}$ vector calculated in Eq. (17.6.4), the corresponding $\underline{W}$ vector from the matrix calculated in Eq. (17.6.5). This is given in matrix notation on the right of Eq. (17.6.6) as the major product of the $\underline{Y}$ vector of Eq. (17.6.4) by a unit row vector less the matrix $\underline{W}$ of Eq. (17.6.5).

We now normalize the column vectors of $\underline{U}$ calculated in Eq. (17.6.6), as indicated in Eq. (17.6.7). We call this the $\underline{V}$ matrix. The right side of Eq.

(17.6.7) shows the $\underline{U}$ matrix of Eq. (17.6.6) postmultiplied by the reciprocal square root of the diagonal of the minor product moment of $\underline{U}$. It is, of course, not necessary to calculate the entire product moment of $\underline{U}$, but only the sums of squares of column elements, in order to get the $\underline{D}$ matrix used in Eq. (17.6.7).

In Eq. (17.6.8) we calculate the matrix $\underline{Z}$, which is the $\underline{V}$ matrix of Eq. (17.6.7) postmultiplied by a diagonal matrix. Now the diagonal matrix is made up of the elements of the diagonal of the minor product moment of the $\underline{V}$ matrix of Eq. (17.6.7) and the $\underline{W}$ matrix of Eq. (17.6.5). Here again, it is not necessary to calculate the minor product moment but only the diagonal elements consisting of the minor products of corresponding columns of the $\underline{V}$ and the $\underline{W}$ matrices.

The next step is indicated by Eq. (17.6.9). This is the matrix $\underline{W}$ of Eq. (17.6.5) minus the $\underline{Z}$ matrix of Eq. (17.6.8). This we indicate as the $\underline{C}$ matrix.

Next we calculate the matrix $\underline{M}$ indicated in Eq. (17.6.10). This is obtained by premultiplying the $\underline{C}$ matrix of Eq. (17.6.9) by the inverse of the upper triangular factor of the matrix $\underline{S}$ in Eq. (17.6.2).

We then calculate the minor product moment $\underline{\gamma}$ of the matrix $\underline{M}$ calculated in Eq. (17.6.10), as indicated in Eq. (17.6.11).

The next step is to calculate a diagonal matrix $\underline{D}$, whose elements are the reciprocal square roots of the diagonal elements of $\underline{\gamma}$ calculated in Eq. (17.6.11).

We calculate the transformation matrix $\underline{H}$ as indicated in Eq. (17.6.13). This consists in normalizing the elements of the $\underline{M}$ matrix of Eq. (17.6.10) as shown on the right hand side of Eq. (17.6.13). The $\underline{M}$ matrix is post-multiplied by the reciprocal square root of the diagonal elements of its minor product moment.

The calculations for the correlations among the primary reference axes are indicated in Eq. (17.6.14). Here we pre- and postmultiply the $\underline{\gamma}$ matrix of Eq. (17.6.11) by the $\underline{D}$ matrix of Eq. (17.6.12).

Finally, we calculate the simple structure factor loading matrix as in Eq. (17.6.15). As in previous methods, we postmultiply the arbitrary factor loading matrix by the transformation matrix $\underline{H}$ calculated in Eq. (17.6.13).

17.6.4 Numerical Example

We use the same group centroid factor matrix as in the preceding section.

The tables for this example will not be discussed in detail. We merely give below the table number and the corresponding equation number where such an equation is given.

Table No.	Equation No.
17.6.1	(17.6.1)
17.6.2	(17.6.2)
17.6.3	--------
17.6.4	(17.6.5)
17.6.5	(17.6.6)
17.6.6	(17.6.7)
17.6.7	(17.6.8)
17.6.8	(17.6.9)
17.6.9	(17.6.10)
17.6.10	(17.6.11)
17.6.11	(17.6.13)
17.6.12	(17.6.15)

It is of interest to note that the elements for each vector of Table 17.6.12 corresponding to the zero element in the corresponding binary hypothesis vector, actually do sum to 0.

Table 17.6.1 - Matrix $\underline{S} = \underline{a' a}$

3.03107	0.58375	0.79188
0.58375	2.38076	0.37794
0.79188	0.37794	1.69911

Table 17.6.2 - Matrix $\underline{t}$

1.74100		
0.33530	1.50610	
0.45484	0.14968	1.21236

Table 17.6.3 - Matrix $\underline{t^{-1}}$

0.57438		
-0.12787	0.66397	
-0.19970	-0.08197	0.82484

Table 17.6.4 - Matrix $\underline{W} = \underline{\alpha' f}$

1.59851	0.26077	0.60885
-0.35654	1.66494	0.21968
-0.55570	-0.30339	1.59539

Table 17.6.5 - Matrix $\underline{U} = \underline{Y\,1'} - \underline{W}$

0.86962	2.20735	1.85928
1.88462	-0.13686	1.30840
1.29200	1.03969	-0.85909

Table 17.6.6 - Matrix $\underline{V} = \underline{U\, D_{U'U}^{-\frac{1}{2}}}$

0.35569	0.90325	0.76501
0.77085	-0.05600	0.53835
0.52846	0.42544	-0.35347

Table 17.6.7 - Matrix $\underline{Z} = \underline{V\, D_{V'W}}$

0.00003	0.01195	0.01538
0.00006	-0.00074	0.01082
0.00004	0.00563	-0.00711

Table 17.6.8 - Matrix $\underline{C} = \underline{W} - \underline{Z}$

1.59848	0.24882	0.59346
-0.35660	1.66568	0.20885
-0.55574	-0.30901	1.60250

Table 17.6.9 - Matrix $\underline{M} = \underline{t'^{-1}\, C}$

1.07472	-0.00836	-0.00586
-0.19121	1.13129	0.00731
-0.45839	-0.25489	1.32180

Table 17.6.10 - Matrix $\underline{\gamma} = \underline{M'\, M}$

1.40172	-0.10847	-0.61359
-0.10847	1.34485	-0.32859
-0.61359	-0.32859	1.74723

Table 17.6.11 - Matrix $\underline{H} = \underline{M\, D_{\gamma}^{-\frac{1}{2}}}$

0.90775	-0.00721	-0.00443
-0.16151	0.97552	0.00553
-0.38718	-0.21979	0.99997

Table 17.6.12 - Matrix $\underline{b} = \underline{a\, H}$

0.857	-0.074	-0.003
0.845	-0.061	0.029
0.825	0.113	-0.038
-0.002	0.802	0.077
-0.028	0.873	-0.042
0.023	0.853	-0.023
-0.028	-0.012	0.776

17.7 The Orthogonal Transformation Matrix

17.7.1 Characteristics of the Method. As indicated earlier in this chapter, it is sometimes desirable to impose the condition of orthonormality on the simple structure transformation matrix. This restriction can be used on any type of arbitrary factor loading matrix which is to be transformed to a simple structure hypothesis matrix. The orthonormal restriction has not been generally used, however. Nevertheless, it has a clear advantage over oblique methods when applied to the principal axis factor loading matrix for which a simple structure binary hypothesis matrix is available. Then the simple structure factor score matrix whose solution we shall consider in a later chapter is an orthonormal matrix. This means that the factor scores are uncorrelated in both the unrotated and the rotated solutions.

One characteristic of this method is that it is much more laborious computationally,and therefore not recommended for desk computers. The method requires successive solutions of the basic structure of certain matrices which are required in the repeated approximations to the final orthonormal transformation matrix.

17.7.2 Computational Equations

17.7.2a Definition of Notation

$\underline{a}$ is an arbitrary factor matrix.

$\underline{f}$ is the binary hypothesis matrix.

$\underline{r}$ is the correlations of the simple structure reference axes.

$\underline{H}$ is the simple structure transformation matrix.

$\underline{b}$ is the simple structure factor matrix.

17.7.2b The Equations

$$\gamma = a' f \tag{17.7.1}$$

$$G = \gamma D^{-1}_{f'f} \tag{17.7.2}$$

$$C_i = a' b_{i-1} - G D_{b'_{i-1}f} \tag{17.7.3}$$

$$C'_i C_i = Q_i \Delta^2_C Q'_i \tag{17.7.4}$$

$$H_i = C_i Q_i \Delta^{-1}_i Q'_i \tag{17.7.5}$$

$$b_i = a H_i \tag{17.7.6}$$

$$r = H'_i H_i = I \tag{17.7.7}$$

17.7.3 Computational Instructions. The computational instructions for this method are brief, but the actual computations can be lengthy even on high speed computers if the number of variables and the number of factors are moderately large, such as those encountered in experimental investigations.

The method consists in a set of successive approximation cycles beginning with any arbitrary factor loading matrix which may or may not be normalized by rows. We have given a binary hypothesis matrix, as in the other methods.

We begin with Eq. (17.7.1), which gives the matrix $\underline{\gamma}$ as the product of the binary matrix premultiplied by the transpose of the factor loading matrix.

We then define a scaling of the $\underline{\gamma}$ matrix as in Eq. (17.7.2). The scaling diagonal is simply the diagonal matrix of the minor product moment of the hypothesis matrix. If this happens to be one in which there is only a single 1 in each row, then each diagonal element is the reciprocal of the number of 1's in the corresponding column of the binary matrix.

We next indicate an iteration cycle by Eq. (17.7.3). Here we have on the right two terms. The first of these includes the preceding approximation to the simple structure factor loading matrix. This first term is an approximation to the simple structure factor matrix premultiplied by the transpose of the arbitrary matrix. When $\underline{i}$ = 1 the approximation to the simple structure matrix, $\underline{b}_0$, may be taken as the simple structure matrix arrived at by any one of the four preceding methods. The second term on the right is the $\underline{G}$ matrix calculated in Eq. (17.7.2), postmultiplied by a diagonal matrix whose elements are the diagonal elements of the minor product of the previous approximation to the simple structure factor matrix and the hypothesis matrix.

We indicate the minor product moment of the matrix calculated in Eq. (17.7.3) by Eq. (17.7.4). Eq. (17.7.4) also indicates the basic structure solution for this minor product moment. For each approximation $\underline{i}$, we calculate all the vectors of the basic structure factors $\underline{\Delta}^2$ and $\underline{Q}$.

Eq. (17.7.5) indicates the $\underline{i}$th approximation to the orthonormal transformation matrix. This is obtained, as indicated on the right, by multiplying from left to right as follows: the $\underline{C}$ matrix by the $\underline{Q}$ matrix by the inverse of the $\underline{\Delta}$ matrix by the transpose of the $\underline{Q}$ matrix.

Finally, we indicate the $\underline{i}$th approximation to the simple structure factor loading matrix by Eq. (17.7.6). This is the arbitrary factor matrix postmultiplied by the $\underline{i}$th approximation to the transformation matrix solved for in Eq. (17.7.5).

This iteration procedure continues until, according to some criterion, the approximations are close enough. Presumably, the trace of the matrix given in Eq. (17.7.4) would provide a satisfactory criterion. When this trace does not change by more than a specified amount from one approximation to another, we may discontinue the computations.

Eq. (17.7.7) assumes that the computations have stabilized, and therefore we have the minor product moment of the current $\underline{H}$ or transformation matrix. This will be a check on the computations and by definition this should be an identity matrix.

17.7.4 Numerical Example. We begin with the same group centroid matrix and binary hypothesis matrix as in the previous section.

Tables 17.7.1 and 17.7.2 show the computations indicated by Eqs. (17.7.1) and (17.7.2), respectively.

The remaining tables are for the 10th approximation, as follows:

Tables 17.7.3 and 17.7.4 give the computation indicated by Eq. (17.7.3).

Table 17.7.5 gives the minor product moment of the matrix $\underline{C}$ in Table 17.7.4.

Table 17.7.6 gives the matrix $\underline{(C'\ C)^{-\frac{1}{2}}} = \underline{Q\,\Delta^{-1}\,Q'}$ which is required in Table 17.7.7 and Eq. (17.7.5).

Table 17.7.7 gives the approximation to the orthogonal transformation matrix $\underline{H}$ as indicated by Eq. (17.7.5).

Table 17.7.8 is the approximation to the simple structure factor matrix, as indicated by Eq. (17.7.6).

Table 17.7.1 - Matrix $\underline{\gamma} = \underline{a' f}$

2.78300	0.45400	1.06000
-0.00100	2.59500	0.53500
0.00000	0.00000	2.24400

Table 17.7.2 - Matrix $\underline{G} = \underline{\gamma D^{-1}_{f'f}}$

0.92767	0.15133	0.35333
-0.00033	0.86500	0.17833
0.00000	0.00000	0.74800

Table 17.7.3 - Matrix $\underline{a' b_{i-1}}$ for $\underline{i} = 10$

0.62649	2.03184	-2.37365
2.39956	0.37041	-0.50667
0.56197	1.82739	0.04159

Table 17.7.4 - Vector $\underline{D^{-1}_{b'_{i-1}f}}$ and Matrix

$$\underline{C_i} = \underline{a' b_{i-1}} - \underline{G D_{b'_{i-1}f}} \text{ for } \underline{i} = 10$$

-0.043	-0.044	0.040

0.66639	2.03846	-2.38776
2.39955	0.40826	-0.51379
0.56197	1.82739	0.01172

Table 17.7.5 - Matrix $\underline{C'_i C_i}$ for $\underline{i} = 10$

6.51772	3.36497	-2.81748
3.36497	7.66132	-5.05569
-2.81748	-5.05569	5.96551

Table 17.7.6 - Matrix $\underline{Q_i \Delta^{-1}_i Q'_i}$ for $\underline{i} = 10$

0.44064	-0.07679	0.06205
-0.07679	0.50770	0.22263
0.06205	0.22263	0.58089

Table 17.7.7 - Matrix $\underline{H_i} = \underline{C_i\ Q_i\ \Delta_i^{-1}\ Q_i'}$ for $\underline{i} = 10$

-0.01105	0.45218	-0.89184
0.99410	-0.09137	-0.05867
0.10803	0.88723	0.44851

Table 17.7.8 - Simple Structure Approximation

$\underline{b_i} = \underline{a\ H_i}$ for $\underline{i} = 10$

-0.078	0.430	-0.827
-0.054	0.457	-0.817
0.100	0.372	-0.838
0.841	0.069	-0.175
0.874	-0.073	-0.168
0.860	-0.027	-0.214
0.244	0.824	0.044
0.169	0.852	0.057
0.349	0.745	-0.072

17.8 Mathematical Proofs

17.8.1 The Multiple Group Matrix

Given the correlation matrix $\underline{R}$ and the binary simple structure hypothesis matrix $\underline{f}$, let

$$F = R f \tag{17.8.1}$$

and

$$S = F' f \tag{17.8.2}$$

$$t t' = S \tag{17.8.3}$$

Then the multiple group factor matrix is well known to be

$$a = F t'^{-1} \tag{17.8.4}$$

Assume now we wish to find the simple structure matrix $\underline{b}$ of best fit to $\underline{f}$. We consider

$$a H = b \tag{17.8.5}$$

and

$$b - f D = \epsilon \tag{17.8.6}$$

where $\underline{D}$ is diagonal and where for $\underline{H}$ in Eq. (17.8.5) we have

$$D_{H'H} = I \tag{17.8.7}$$

We wish to minimize

$$\mathrm{tr}\ \epsilon'\epsilon = \psi \qquad (17.8.8)$$

The solution is well known to be

$$H = (a'\ a)^{-1}\ a'\ f\ D \qquad (17.8.9)$$

From Eqs. (17.8.2), (17.8.3), (17.8.4), and (17.8.9)

$$H = t'\ (F'\ F)^{-1}\ S\ D \qquad (17.8.10)$$

Let

$$F'\ F = G \qquad (17.8.11)$$

From Eqs. (17.8.3), (17.8.10), and (17.8.11)

$$H'\ H = D\ S\ G^{-1}\ S\ G^{-1}\ S\ D \qquad (17.8.12)$$

Let

$$C = G^{-1}\ S \qquad (17.8.13)$$

and

$$g = S\ C \qquad (17.8.14)$$

From Eqs. (17.8.7), (17.8.12), (17.8.13), and (17.8.14)

$$H'\ H = D\ g\ C\ D \qquad (17.8.15)$$

From Eq. (17.8.15)

$$D = D_g^{-\frac{1}{2}}\ C \qquad (17.8.16)$$

From Eqs. (17.8.4), (17.8.5), (17.8.11), and (17.8.16)

$$b = F\,C\,D_{g\,C}^{-\frac{1}{2}} \qquad (17.8.17)$$

17.8.2 The Principal Axis Matrix

Given the principal axis factor matrix

$$a = Q\,\delta^{\frac{1}{2}} \qquad (17.8.18)$$

and the binary hypothesis matrix $\underline{f}$, consider the least square transformation

$$a\,H = b \qquad (17.8.19)$$

such that in

$$b - f\,D = \epsilon \qquad (17.8.20)$$

The trace of $\underline{\epsilon'\epsilon}$ is minimized with a diagonal $\underline{D}$ such that for $\underline{H}$ in Eq. (17.8.19) we have

$$D_{H'\,H} = I \qquad (17.8.21)$$

From Eqs. (17.8.19) and (17.8.20)

$$a\,H - f\,D = \epsilon \qquad (17.8.22)$$

The solution for $\underline{H}$ is obviously

$$H = (a'\,a)^{-1}\,a'\,f\,D \qquad (17.8.23)$$

From Eq. (17.8.18) in Eq. (17.8.23)

$$H = \delta^{-1}\,a'\,f\,D \qquad (17.8.24)$$

From Eq. (17.8.24)

$$H' H = D (f' a) \delta^{-2} (a' f) D \qquad (17.8.25)$$

Let

$$a' f = G \qquad (17.8.26)$$

$$\delta^{-1} G = C \qquad (17.8.27)$$

From Eqs. (17.8.25), (17.8.26), and (17.8.27)

$$H' H = D C' C D \qquad (17.8.28)$$

From Eqs. (17.8.21) and (17.8.28)

$$D = D_{C'C}^{-\frac{1}{2}} \qquad (17.8.29)$$

From Eqs. (17.8.24), (17.8.26), (17.8.27), and (17.8.29)

$$H = C D \qquad (17.8.30)$$

17.8.3 The Arbitrary Matrix

Let $\underline{a}$ be any factor loading matrix, $\underline{f}$ the binary hypothesis matrix, and consider

$$A = D_{a a'}^{-\frac{1}{2}} a \qquad (17.8.31)$$

so that the rows of $\underline{A}$ are normalized.

Consider

$$A H = b \qquad (17.8.32)$$

and

$$b - f D = \epsilon \tag{17.8.33}$$

with $\underline{H}$ determined so that $\underline{tr}\ \underline{\epsilon'\epsilon}$ is minimized and $\underline{D}$ is a diagonal such that

$$D_{H'H} = I \tag{17.8.34}$$

Then the solution for $\underline{H}$ is well known to be

$$H = (A'\ A)^{-1} A'\ f\ D \tag{17.8.35}$$

Let

$$M = (A'\ A)^{-1} A'\ f \tag{17.8.36}$$

From Eqs. (17.8.34) and (17.8.36)

$$D = D_{M'M}^{-\frac{1}{2}} \tag{17.8.37}$$

and from Eqs. (17.8.35), (17.8.36), and (17.8.37)

$$H = M\ D \tag{17.8.38}$$

17.8.4 The Zero Partial Sum Simple Structure Matrix

Given the arbitrary factor matrix $\underline{a}$ and the binary hypothesis matrix $\underline{f}$, let

$$L = 1\ 1' - f \tag{17.8.39}$$

be called the supplementary matrix to $\underline{f}$. Consider

$$a\ M = b \tag{17.8.40}$$

and

$$b - f = \epsilon \tag{17.8.41}$$

with the restriction that

$$D_{b'L} = 0 \tag{17.8.42}$$

That is, the hypothesized 0's in $\underline{f}$ sum to 0 in $\underline{b}$ for corresponding columns. From Eqs. (17.8.40), (17.8.41), and (17.8.42) we write

$$\mathrm{tr}\,(\epsilon' \epsilon - 2\, D_{b'L}\, D_\lambda) = \psi \tag{17.8.43}$$

where $\underline{D_\lambda}$ is a diagonal matrix of Lagrangian multipliers.

It will be simpler now to consider Eq. (17.8.43) with respect to each $\underline{M_{.i}}$ vector separately. From Eqs. (17.8.40), (17.8.41), and (17.8.43)

$$M'_{.i}\, a'\, a\, M_{.i} - 2\, M'_{.i}\, a'\, f_{.i} + f'_{.i}\, f_{.i} - 2\, M'_{.i}\, a'\, L_{.i}\, \lambda_i = \psi_i \tag{17.8.44}$$

Differentiating Eq. (17.8.45) symbolically with respect to $\underline{M'_{.i}}$ and equating to 0 gives

$$\frac{\partial \psi_i}{\partial M'_{.i}} = 2\,(a'\, a\, M_{.i} - a'\, f_{.i} - a'\, L_{.i}\, \lambda_i) = 0 \tag{17.8.45}$$

From Eq. (17.8.45)

$$M_{.i} = (a'\, a)^{-1}\, a'\, f_{.i} + (a'\, a)^{-1}\, a'\, L_{.i}\, \lambda_i \tag{17.8.46}$$

Premultiplying Eq. (17.8.46) by $\underline{L'_{.i}\, a}$ and using Eqs. (17.8.40) and (17.8.42)

$$\lambda_i = - \frac{L'_{.i}\ a\ (a'\ a)^{-1}\ a'\ f_{.i}}{L'_{.i}\ a\ (a'\ a)^{-1}\ a'\ L_{.i}} \tag{17.8.47}$$

Let

$$t\ t' = a'\ a \tag{17.8.48}$$

$$\alpha = a\ t'^{-1} \tag{17.8.49}$$

Using Eqs. (17.8.48) and (17.8.49) in Eqs. (17.8.46) and (17.8.47), we get, respectively,

$$M_{.i} = t'^{-1}\ (\alpha'\ f_{.i} - a'\ L_{.i}\ \lambda_i) \tag{17.8.50}$$

$$\lambda_i = - \frac{L'_{.i}\ \alpha\ \alpha'\ f_{.i}}{L'_{.i}\ \alpha\ \alpha'\ L_{.i}} \tag{17.8.51}$$

Using Eq. (17.8.51) in Eq. (17.8.50)

$$M_{.i} = t'^{-1}\ (\alpha'\ f_{.i} - \frac{\alpha'\ L_{.i}\ L'_{.i}\ \alpha\ \alpha'\ f_{.i}}{L'_{.i}\ \alpha\ \alpha'\ L_{.i}}) \tag{17.8.52}$$

or

$$M_{.i} = t'^{-1}\ (I - \frac{\alpha'\ L_{.i}\ L'_{.i}\ \alpha}{L'_{.i}\ \alpha\ \alpha'\ L_{.i}})\ \alpha'\ f_{.i} \tag{17.8.53}$$

Let

$$\alpha'\ L_{.i} = U_{.i} \tag{17.8.54}$$

$$V_{.i} = \frac{U_i}{\sqrt{U'_{.i}\ U_{.i}}} \tag{17.8.55}$$

$$\alpha' \, f_{.i} = W_{.i} \tag{17.8.56}$$

From Eqs. (17.8.54), (17.8.55), and (17.8.56) in Eq. (17.8.53)

$$M_{.i} = t'^{-1} \, (I - V_{.i} \, V'_{.i}) \, W_{.i} \tag{17.8.57}$$

Going now to the complete matrix notation, we have from Eqs. (17.8.54), (17.8.55), and (17.8.56), respectively,

$$\alpha' \, L = U \tag{17.8.58}$$

$$V = U \, D^{-\frac{1}{2}}_{U'U} \tag{17.8.59}$$

$$\alpha' \, f = W \tag{17.8.60}$$

From Eqs. (17.8.39) and (17.8.58)

$$U = \alpha' \, (1 \, 1' - f) \tag{17.8.61}$$

From Eqs. (17.8.60) and (17.8.61)

$$U = \alpha' \, 1 \, 1' - W \tag{17.8.62}$$

Or, if we let

$$\alpha' \, 1 = Y \tag{17.8.63}$$

we have from Eq. (17.8.63) in Eq. (17.8.62)

$$U = Y \, 1' - W \tag{17.8.64}$$

Using Eqs. (17.8.58), (17.8.59), and (17.8.60) in Eq. (17.8.57)

$$M = t'^{-1} (W - V D_{V'W}) \tag{17.8.65}$$

If now we wish to normalize $\underline{M}$ we have

$$H = M D_{M'M}^{-\frac{1}{2}} \tag{17.8.66}$$

17.8.5 The Orthogonal Transformation Matrix

Suppose we have a binary hypothesis matrix $\underline{f}$, and an arbitrary factor matrix $\underline{a}$ which we wish to transform by a square orthonormal transformation to the best least square approximation to $\underline{f\ D}$, where $\underline{D}$ is a diagonal to be determined. We let

$$b = a\ H \tag{17.8.67}$$

where by hypothesis

$$H'\ H = I \tag{17.8.68}$$

The approximation equation is

$$b - f\ D = \epsilon \tag{17.8.69}$$

We write the least square function with the constraint in Eq. (17.8.69) as

$$\mathrm{tr}\ (\epsilon'\ \epsilon - H'\ H\ \lambda) = \psi \tag{17.8.70}$$

where $\underline{\lambda}$ is a matrix of Lagrangian multipliers and

$$\lambda = \lambda' \tag{17.8.71}$$

From Eqs. (17.8.67) and (17.8.69) in Eq. (17.8.70), we have

$$\psi = \mathrm{tr}\,(H'\,a'\,a\,H - 2\,H'\,a'\,f\,D + D\,f'\,f\,D - H'\,H\,\lambda) \qquad (17.8.72)$$

Differentiating Eq. (17.8.71) symbolically with respect to $\underline{H'}$ and equating to 0 gives

$$\frac{\partial \psi}{\partial H'} = 2\,(a'\,a\,H - a'\,f\,D - H\,\lambda) = 0 \qquad (17.8.73)$$

To differentiate Eq. (17.8.71) with respect to $\underline{D}$, we let

$$V_D = D\,1 \qquad (17.8.74)$$

Using Eqs. (17.8.67) and (17.8.74) in Eq. (17.8.72) gives

$$\psi = \mathrm{tr}\,(b'\,b - 2\,V_D'\,D_{b'f}\,1 + V_D'\,D_{f'f}\,V_D - H'\,H\,\lambda) \qquad (17.8.75)$$

Differentiating Eq. (17.8.75) symbolically with respect to $\underline{V_D'}$ and equating to 0 gives

$$\frac{\partial \psi}{\partial V_D'} = -2\,(D_{b'f}\,1 - D_{f'f}\,V_D) = 0 \qquad (17.8.76)$$

From Eq. (17.8.76)

$$D = D_{f'f}^{-1}\,D_{b'}\,f \qquad (17.8.77)$$

Using Eqs. (17.8.67) and (17.8.77) in Eq. (17.8.73)

$$a'\,b - a'\,f\,D_{f'f}^{-1}\,D_{b'f} - H\,\lambda = 0 \qquad (17.8.78)$$

Let

$$a' \, f \, D^{-1}_{f'f} \; = \; G \qquad (17.8.79)$$

From Eqs. (17.8.78) and (17.8.79)

$$(a' \, b - G \, D_{b'f}) \, \lambda^{-1} \; = \; H \qquad (17.8.80)$$

Let

$$a' \, b - G \, D_{b'f} \; = \; C \qquad (17.8.81)$$

and

$$C \; = \; P \, \Delta \, Q' \qquad (17.8.82)$$

From Eqs. (17.8.81) and (17.8.82) in Eq. (17.8.80)

$$P \, \Delta \, Q' \, \lambda^{-1} \; = \; H \qquad (17.8.83)$$

Because of Eqs. (17.8.68) and (17.8.72), the only $\underline{\lambda^{-1}}$ which will satisfy Eq. (17.8.83) is

$$\lambda^{-1} \; = \; Q \, \Delta^{-1} \, Q' \qquad (17.8.84)$$

Therefore, from Eqs. (17.8.83) and (17.8.84)

$$H \; = \; P \, Q' \qquad (17.8.85)$$

From Eq. (17.8.82) we have

$$C' \, C \; = \; Q \, \Delta^2 \, Q' \qquad (17.8.86)$$

From Eq. (17.8.82)

$$P \; = \; C \, Q \, \Delta^{-1} \qquad (17.8.87)$$

From Eqs. (17.8.85) and (17.8.87)

$$H = C\, Q\, \Delta^{-1}\, Q' \tag{17.8.88}$$

From Eqs. (17.8.67) and (17.8.88) we have

$$b = a\, C\, Q\, \Delta^{-1}\, Q' \tag{17.8.89}$$

We may start with the approximation

$$b_0 = a \tag{17.8.90}$$

Then

$$C_1 = a'\, a - G\, D_{a'f} \tag{17.8.91}$$

$$C_1'\, C_1 = Q_1\, \Delta_1^2\, Q_1' \tag{17.8.92}$$

$$H_1 = C_1\, Q_1\, \Delta_1^{-1}\, Q_1' \tag{17.8.93}$$

In general,

$$C_i = a'\, b_{i-1} - G\, D_{b_{i-1}'f} \tag{17.8.94}$$

$$C_i'\, C_i = Q_i\, \Delta_i^2\, Q_i' \tag{17.8.95}$$

$$H_i = C_i\, Q_i\, \Delta_i^{-1}\, Q_i' \tag{17.8.96}$$

$$b_i = a\, H_i \tag{17.8.97}$$

CHAPTER 18

ANALYTICAL ROTATIONS

We saw in Chapter 17 that if we have some hypothesis as to which variables have high loadings and which have low loadings in each factor, we may set up a binary hypothesis matrix and, by least square procedures with certain constraints on the transformation, get the best approximation to the binary matrix. We saw that the transformation matrices for these procedures are not, in general, square orthonormal unless we impose this additional constraint, as in the last method of that chapter. In many cases, however, a binary matrix may not be available, or it may be that the hypotheses are poorly satisfied by the data.

It is therefore desirable to have analytical methods which are independent of the *a priori* hypotheses of the experimenter. These analytical methods presumably should approximate the criteria outlined in the introduction to Chapter 17. These are the criteria formulated by Thurstone (1947).

A great many methods have been proposed for analytical rotations to simple structure factor loading matrices. The earlier methods were proposed by Thurstone (1947), followed by several methods developed by Horst (1941) and Tucker (1944) which were semi-analytical. Later Wrigley and Newhouse (1952) proposed more completely analytical procedures. Then followed the work of Carroll (1953), Saunders (1953), and several others. Perhaps the best known methods are based on the work of Kaiser (1958).

In any case, although the mathematical thinking and development which have gone into many of the proposed analytical methods is ingenious, these methods have not resulted in the success which may have been hoped for. The methods in general are very laborious computationally, even with the high speed computers, and often they do not give results which come close to satisfying the criteria of simple structure.

At least two conditions should be satisfied by analytical methods, or by any method of rotation. First, the factor loadings should be relatively invariant with respect to the group of entities on which the data are collected. Second, a subset of factor loadings should be relatively invariant, irrespective of which particular battery of variables includes that subset. This latter criterion is subject to certain further qualifications, but one criterion of a good transformation procedure is that the factor loadings of variables be relatively invariant, both with respect to the sampling of entities and the sampling of attributes.

One of the chief difficulties with most of the analytical methods which have been developed is that they are greatly influenced by the particular selection of variables which go into the correlation matrix. In this book we shall not attempt to give an account of all of the analytical methods which have been proposed. We shall, however, briefly describe the methods of Professor John Carroll, whose pioneering and ingenious work may eventually result in more adequate methods.

Carroll (1953) proposed that we have a minimum number of negative factor loadings, and that such as were present should be small. His criterion for transformation was based on the squared factor loadings of the transformed matrix. Therefore the signs would not influence the criterion of goodness of transformation. For this matrix he required in his early model that the minor products of all possible pairs of vectors be a minimum. This meant, in effect, that for any pair there should be a number of very small squared factor loadings; there should be very few factor loadings which were high for both factors; and that, for those which were high in one factor, the loadings should be low in the other, and vice versa. Here, then, are included three of the Thurstone criteria of simple structure. These conditions would, of course, satisfy the criterion that the minor product moment of the two vectors of squared factor loadings should be small.

Carroll (1953) worked out an ingenious computational procedure for the high speed computer for achieving such a minimum for all pairs of factor loading vectors. The difficulties with the procedure were, first, it was strongly influenced by the particular variables in the set, and second, it resulted in too few high factor loadings and too many negative loadings of medium size. The transformation matrix in general was such that the correlations of the reference axes were negative, and the correlation among the primary factors, as discussed in Chapter 21, tended to be positive.

Later, Carroll (1957) changed this criterion by considering, not the minimization of sums of minor products for all pairs of squared element factor loading vectors, but rather the minimization of the covariance of these vectors. This resulted in an overcorrection for the limitations of the previous methods.

The earlier of these methods was called the quartimin method, and the later one was called the covarimin method. Carroll found from empirical investigation that neither of these methods worked very well. The former procedure was biased in favor of reference axes, which were too low in correlation among themselves, and the latter method resulted in reference axes which were too highly correlated. For the latter method, the large simple structure factor loadings tended to be too large and the small ones tended to be considerably greater than 0.

Carroll then formulated a combination of the two methods so as to neutralize the undesirable effects of both. This combination of the quartimin and covarimin methods has been called the oblimin method. The procedure still left a decision as to just how to combine the two procedures, and a certain amount of arbitrariness remained.

The great advantage of Carroll's approach is that one need not hypothesize as to whether a transformation is orthonormal or oblique. The solution itself purports to solve for the correlations among the simple structure reference axes and the primary factor axes. Unfortunately, even with the ingenious rationale and the extraordinarily elaborate computational procedures which have been worked out, the methods still have not demonstrated their usefulness for some sets of experimental data.

Currently, it appears that the work of Kaiser (1958) has had more practical impact on the work of factor analysts than that of other investigators. The procedures of Kaiser specify an orthonormal transformation. This makes the mathematics and the computational routines considerably more straightforward and amenable to the application of the basic structure concepts. It does impose limitations on the results to be expected. In particular, the possibility of achieving relative invariance of transformed factor loadings with respect to both sample of entities and sample of attributes is more remote than if more general transformation procedures were available. However, since in this book our emphasis is on practical application, we shall give our major attention to the methods developed by Kaiser, and variations of these.

18.1 Characteristics of the Methods

18.1.1 The Orthonormal Transformation Matrix. The methods we shall consider do not admit of oblique transformations. The mathematical models on which they are based, and the computational routines, have the restriction of orthonormality built into them. While even yet a number of investigators prefer the orthonormal type of transformation as more desirable from a philosophical and scientific point of view, the tendency seems to be gaining ground to prefer the earlier objectives of Thurstone in his relaxed oblique transformations, which he was able to achieve by graphical methods and shrewd subjective judgment. Unfortunately, many of his followers were not able to apply the same ingenious insights and judgments in their efforts to use the non-analytical graphical methods.

It is probable that the quest to relax the orthonormal transformation by satisfactory objective analytical procedures will eventually triumph. If so, it will probably result in methods for objectively eliminating from the variables contributing to the simple structure determination, those which are most complex in structure and which tend to confuse the transformation attempts.

18.1.2 The Optimizing Function. With the constraint that the transformation shall be orthonormal, the class of solutions we shall discuss all consider a transformed matrix whose elements have been raised to some even power. This means that the new matrix has all positive elements.

In particular, we may consider the matrix of squared elements of transformed factor loadings, as did Carroll. Our attention, however, is directed to only a single factor vector of these positive elements at a time. Specifically, the optimizing criterion which Kaiser (1958) suggested is that the variance of each such vector of positive elements shall be a maximum.

Since the elements are all positive by hypothesis, the maximum for each vector would be achieved if all of its elements were either large or 0. Therefore, in working toward the maximization of this variance criterion, one tends to reduce the intermediate loadings to a minimum and to maximize the number of large and small loadings. This again satisfies one of the criteria of Thurstone for simple structure, i.e., that each factor should have a relatively large number of near vanishing elements, a limited number of very large elements, with very few elements of intermediate size.

18.1.3 Iterative Type Solutions. All of the models considered in this chapter differ essentially from most of those in Chapter 17 in that the solution for the orthonormal transformation matrix is arrived at by successive approximations. It will be recalled that in the last chapter only the last model required successive approximations. This is the one in which the restriction of orthonormality of the transformation is imposed.

The varimax solution, as it was called by Kaiser and as developed by him, consists of a large number of orthonormal transformations involving only two factor vectors at a time. The procedure in general is to start with, say, the first two factor vectors, and transform them by an orthonormal transformation so that the variance of their squared elements is a maximum. One then proceeds with the new first and the third vectors, and applies another orthonormal transformation which satisfies the criterion of maximum variance of squared elements for the transformed vectors. This procedure continues for all possible pairs until the criterion of variances of squared factor loadings ceases to improve.

It can be seen that this would be an extremely laborious procedure for desk calculators. Even for the high speed computers it can be expensive and time-consuming if the matrices to be transformed are very large--for example, of the order of 500 attributes by 20 or 30 factors. After the varimax criterion is satisfied, the method of Kaiser requires the product of all of the orthonormal matrices involved in the computational routine. This cumulative product gives the orthonormal transformation which, when applied directly to the arbitrary factor loading matrix, yields a transformed matrix satisfying the varimax criterion.

18.1.4 Accumulation of Decimal Error. The method of Kaiser, because of the very large number of individual computations going into the procedure, each of which involves rounding errors, is subject to the accumulation of considerable decimal error if the number of variables is large. The methods we shall outline use somewhat different approaches to achieve the varimax criterion. They do not in general accumulate as much decimal error as those of Kaiser. As a matter of fact, they are self-correcting with respect to both decimal and computational error.

18.2 Kinds of Methods

We shall discuss four variations of the methods proposed by Kaiser. These we may call the successive factor varimax, the simultaneous factor varimax, the successive factor general varimax, and the simultaneous factor general varimax.

18.2.1 Successive Factor Varimax. The successive factor varimax method differs essentially from that of Kaiser in that we solve for one factor vector at a time, rather than for all of them simultaneously. As each factor is solved for, it satisfies the varimax criterion in that the variance of the sums of the squared elements for a factor is a maximum. Having found this factor, we find another in which the transformation vector is orthogonal to the first. With this restriction, the variance of the squared elements of the next factor vector is a maximum. We continue in this way until we have found the last factor.

18.2.2 Simultaneous Factor Varimax. In this model we start with an approximation of some sort to the simple structure matrix of factor loadings. We then solve for a second approximation to the factor loading matrix which will satisfy the varimax criterion. We thus proceed by successive approximations to get factor loading matrices which will yield better and better approximations to the matrix which ultimately best satisfies the varimax criterion.

The restriction for each approximation is always that the transformation matrix for that approximation is orthonormal. The procedure therefore yields all of the final simple structure factor loading vectors simultaneously, rather than one at a time.

18.2.3 The Successive Factor General Varimax. This method is like the successive factor varimax, which gets one factor at a time and maximizes the variances of the squared factor loadings, except that we require that some even power of the factor loading elements be positive. This even power may be any value not less than unity. In particular, we can require that the variance of the absolute values of the factor loadings be maximum.

18.2.4 Simultaneous Factor General Varimax. This model is similar to the one discussed in Section 18.2.2 in that, by successive iterations, the simple structure factor loading matrix is solved for by approximations to all of the factor vectors at one time. It is similar to the model discussed in Section 18.2.3 in that the criterion which is maximized is a generalization of the variance of the squared factor loadings. Here again, we maximize the variance of some positive even power of the factor loadings, where the power is not less than unity.

As will be seen in the mathematical proof, Section 18.7.4, one may take a positive even power which in the limiting case approaches the absolute value of the factor loadings. On the other hand, one could maximize the variance of the fourth powers, or the four-thirds powers, or any other powers in which the numerator of the exponent is even and the denominator odd and less than the numerator. It can be seen that, if the numerator of the exponent is an even number approaching infinity, and the odd number is always one less than the numerator, an element raised to this power would approach the absolute value of the element.

18.3 The Successive Factor Varimax Solution

18.3.1 Characteristics of the Method. It has been indicated that the successive factor varimax solution does not give the same answer as the method of Kaiser, in which the transformations are made two vectors at a time. Actually, in the former case the factor loading vectors for the transformed solution tend to come out in the order of the variance of their squared loadings.

This may not always be the case, as the final result may depend somewhat on the approximation one starts with. There is currently no mathematical proof to indicate whether, or under what conditions, this might be true.

Perhaps the chief advantage of this method is that the simple structure factors tend to come out in the order of clarity of interpretation, so that one may neglect the factors which appear later in the solution if they seem to be too obscure or ambiguous. The method is different in this respect from Kaiser's, since the ambiguity of the simple structure factors for his method seems to be spread over all the factors approximately equally.

18.3.2 Computational Equations

18.3.2a Definition of Notation

$\underline{{}_1a}$ is the arbitrary factor matrix.

$\underline{H}$ is the orthogonal transformation matrix.

$\underline{b}$ is the simple structure factor matrix.

$\underline{b^{(2)}}$ is a matrix whose elements are the second powers of the elements in $\underline{b}$.

$\underline{b^{(3)}}$ is a matrix whose elements are the third powers of the elements in $\underline{b}$.

18.3.2b The Equations

$$W = \frac{a' 1}{\sqrt{(1' a)(a' 1)}} \qquad (18.3.1)$$

$$ {}_1V = {}_1a\, W \qquad (18.3.2)$$

$$ {}_{1}V_{L_1} = \min \qquad (18.3.3)$$

$$ {}_{0}H_{.1} = {}_{1}a_{L_1}. \qquad (18.3.4)$$

$$ {}_{1}b_{.1} = {}_{1}a \; {}_{0}H_{.1} \qquad (18.3.5)$$

$$ {}_{1}\beta_{.1} = {}_{1}b^{(3)}_{.1} - {}_{1}b_{.1} \; \frac{1' \; {}_{1}b^{(2)}_{.1}}{n} \qquad (18.3.6)$$

$$ {}_{1}U_{.1} = {}_{1}a' \; {}_{1}\beta_{.1} \qquad (18.3.7)$$

$$ {}_{1}\alpha_{1} = \sqrt{{}_{1}U'_{.1} \; {}_{1}U_{.1}} \qquad (18.3.8)$$

$$ {}_{1}H_{.1} = \frac{{}_{1}U_{.1}}{{}_{1}\alpha_{1}} \qquad (18.3.9)$$

$$ {}_{2}b_{.1} = {}_{1}a \; {}_{1}H_{.1} \qquad (18.3.10)$$

$$ {}_{s}\beta_{.1} = {}_{s}b^{(3)}_{.1} - {}_{s}b_{.1} \; \frac{1' \; {}_{s}b^{(2)}_{.1}}{n} \qquad (18.3.11)$$

$$ {}_{s}U_{.1} = {}_{1}a' \; {}_{s}\beta_{.1} \qquad (18.3.12)$$

$$ {}_{s}\alpha_{1} = \sqrt{{}_{s}U'_{.1} \; {}_{s}U_{.1}} \qquad (18.3.13)$$

$$ {}_{s}H_{.1} = \frac{{}_{s}U_{.1}}{{}_{s}\alpha_{1}} \qquad (18.3.14)$$

$$ {}_{s+1}b_{.1} = {}_{1}a \; {}_{s}H_{.1} \qquad (18.3.15)$$

$$ {}_2a = {}_1a - b_{.1}\, H'_{.1} \tag{18.3.16} $$

$$ {}_iV = {}_{i-1}V + b_{.i-1} \tag{18.3.17} $$

$$ {}_iV_{L_i} = \min\, {}_iV \tag{18.3.18} $$

$$ {}_ia_{L_i.} = {}_0H_{.i} \tag{18.3.19} $$

$$ {}_1b_{.i} = {}_ia\; {}_0H_{.i} \tag{18.3.20} $$

$$ {}_s\beta_{.i} = {}_sb^{(3)}_{.i} - {}_sb_{.i}\; \frac{1'\; {}_sb^{(2)}_{.i}}{n} \tag{18.3.21} $$

$$ {}_sU_{.i} = {}_ia'\; {}_s\beta_{.i} \tag{18.3.22} $$

$$ {}_s\alpha_i = \sqrt{{}_sU'_{.i}\; {}_sU_{.i}} \tag{18.3.23} $$

$$ {}_sH_{.i} = \frac{{}_sU_{.i}}{{}_s\alpha_i} \tag{18.3.24} $$

$$ {}_{s+1}b_{.i} = {}_ia\; {}_sH_{.i} \tag{18.3.25} $$

$$ {}_{i+1}a = {}_ia - b_{.i}\; H'_{.i} \tag{18.3.26} $$

18.3.3 Computational Instructions. We begin with an arbitrary factor loading matrix ${}_{\underline{1}}a$. Kaiser has recommended, and the practice seems to be generally desirable, that any arbitrary factor loading matrix, before it is transformed or rotated, should be normalized by rows. We shall therefore assume, in this and the succeeding models, that the arbitrary factor loading matrices have been normalized by rows.

The rationale for selecting the first approximation to the first transformation vector is as follows. We assume that if there were, in the set of tests or measures, one which measured one of the primary factors rather accurately, it would have a relatively low correlation with the average of all the variables. Therefore we calculate a normalized vector of the average of the factor loadings by columns, as indicated in Eq. (18.3.1). The right hand side of this equation gives in the numerator a column vector whose elements are the sums of the column elements of the factor loading matrix ${}_{\underline{1}}a$. As can be seen, the denominator scalar of this right hand term is the square root of the minor product moment of the vector in the numerator. Therefore, the $\underline{W}$ column vector on the left of Eq. (18.3.1) is a normal vector.

In Eq. (18.3.2) we calculate the vector ${}_{\underline{1}}V$. This is the factor loading matrix ${}_{\underline{1}}a$ postmultiplied by the vector $\underline{W}$ of Eq. (18.3.1). This now gives a vector of correlations of the average of all the tests with each of the measures. Presumably, that variable which correlates lowest with this average would be a relatively pure measure of a factor.

We therefore look for the lowest element in the vector given by Eq. (18.3.2). This is indicated in Eq. (18.3.3). We use the subscript $L_{\underline{1}}$ to indicate the position of this lowest value.

We then take the $L_{\underline{1}}$ row vector of the ${}_{\underline{1}}a$ factor loading matrix as the zero approximation to the first vector of the transformation matrix $\underline{H}$, as indicated in Eq. (18.3.4).

Next we postmultiply the factor loading matrix by the vector indicated in Eq. (18.3.4). This gives the first approximation to the first transformed factor loading vector, as indicated in Eq. (18.3.5).

Then we calculate the first approximation to a $\underline{\beta}$ vector, as shown in Eq. (18.3.6). On the right hand side of the equation, the first term is a vector whose elements are the cubes of the elements of the vector calculated in Eq. (18.3.5). The second term on the right of Eq. (18.3.6) is the vector calculated in Eq. (18.3.5), multiplied by a scalar quantity which is the average of the sums of squares of the elements of the vector in Eq. (18.3.5). This second vector is subtracted from the first.

Next we calculate the first approximation to the $U_{\underline{.1}}$ vector, as indicated in Eq. (18.3.7). This is the transpose of the factor loading matrix postmultiplied by the vector calculated in Eq. (18.3.6).

We now calculate a scalar quantity as in Eq. (18.3.8). This is the square root of the minor product moment of the vector calculated in Eq. (18.3.7).

Next we calculate the first approximation to the first transformation vector as in Eq. (18.3.9). The vector calculated in Eq. (18.3.7) is divided by the scalar calculated in Eq. (18.3.8). We see, therefore, that the vector calculated in Eq. (18.3.9) is a normal vector.

We now calculate the second approximation to the transformed factor loading vector $b_{\underline{.1}}$. As indicated in Eq. (18.3.10), we postmultiply the arbitrary factor loading matrix by the vector calculated in Eq. (18.3.9).

We continue to calculate successive approximations to the first transformation vector $H_{\underline{.1}}$ and the first simple structure factor vector $b_{\underline{.1}}$, as indicated in Eqs. (18.3.11) through (18.3.15). These equations are the same as Eqs. (18.3.6) through (18.3.10), except that the prescript of 1 has been changed to the general subscript $\underline{s}$, which means the $\underline{s}$ approximation.

The stabilization limit may be based on the scalar α indicated in Eqs. (18.3.8) and (18.3.13). When this scalar has stabilized to a sufficient degree of accuracy, we may assume that the $H_{.1}$ vector is sufficiently accurate, and therefore that the $b_{.1}$ vector is also sufficiently accurate.

We then calculate a residual factor loading matrix, as indicated in Eq. (18.3.16). The first term on the right side is the factor loading matrix a with a prescript 1. We use this prescript to show that it is the original arbitrary factor loading matrix, rather than some residual matrix derived from it. The second term on the right of this equation is the major product moment of the factor loading vector $b_{.1}$ and the transformation vector $H_{.1}$. This major product is subtracted from the factor loading matrix to give a residual matrix ${}_2a$.

We are now ready to begin the computations for the second simple structure factor vector $b_{.2}$, and the second transformation vector $H_{.2}$. We require a first approximation to the $H_{.2}$ vector. This is accomplished by considering Eq. (18.3.17). Here for the subscript i we substitute 2. On the right hand side of the equation this gives as the first term ${}_1V$, which we calculated in Eq. (18.3.2). To this is added $b_{.1}$, calculated in the previous cycle of computations, to give the vector V with a prescript of 2.

We now consider Eq. (18.3.18) in which the i subscript takes the value of 2. This equation means that we find the smallest value in the vector calculated in Eq. (18.3.17) and call this the L_2 position.

Having identified this position, we then take as our zero approximation to the $H_{.2}$ vector the L_2 row of the residual factor loading matrix ${}_2a$ calculated in Eq. (18.3.16). This is indicated in Eq. (18.3.19).

Next we calculate the first approximation to $b_{.2}$ as indicated in Eq. (18.3.20), in which $i = 2$. The right hand side of this equation shows that we postmultiply the matrix calculated in Eq. (18.3.16) by the vector for $i = 2$ from Eq. (18.3.19).

For the computation of the s approximation to the ith transformation vector $H_{.i}$, and the $s+1$ approximation to the ith factor loading vector $b_{.i}$, we now have the series of equations (18.3.21) through (18.3.25).

Eq. (18.3.26) shows the general equation for calculating the $i+1$ residual factor loading matrix ${}_{i+1}a$ from the ith residual factor loading matrix ${}_{i}a$, the ith factor loading vector $b_{.i}$, and the ith transformation vector $H_{.i}$.

18.3.4 Numerical Example. We shall use the same numerical example throughout to illustrate the various models in this chapter.

We begin with the first three factors of the principal axis factor loading matrix which are given in Table 18.3.1.

Table 18.3.2 gives the final matrix H which transforms the arbitrary matrix a to the varimax simple structure matrix b.

Table 18.3.3 shows the final approximation to the varimax factor loading matrix. It can be verified that the rows of this matrix are normalized. If desired, they may be scaled back to the variances of the rows of the principal axis matrix whose transpose is given in Table 18.3.1.

Table 18.3.1 - Transpose of Principal Axis Factor Loading Matrix

0.717	0.740	0.773	0.556	0.463	0.518	0.640	0.615	0.715
0.493	0.478	0.296	-0.649	-0.744	-0.694	0.080	0.166	-0.034
0.350	0.322	0.406	0.068	0.181	0.188	-0.588	-0.621	-0.369

Table 18.3.2 - Final Transformation Matrix H for Successive Factor Varimax Solution

0.502	0.702	0.505
-0.842	0.531	0.098
0.199	0.475	-0.857

Table 18.3.3 - Final Varimax Factor Matrix b for Successive Factor Varimax Solution

0.016	0.993	0.118
0.035	0.987	0.154
0.238	0.968	0.077
0.978	0.091	0.185
1.000	0.018	0.006
0.995	0.095	0.036
0.157	0.244	0.957
0.051	0.253	0.966
0.390	0.384	0.837

18.4 Simultaneous Factor Varimax Solution

18.4.1 Characteristics of the Method. The computational procedures in this method are essentially different from those of the method just described. Successive iterations are required, but we iterate simultaneously to all of the factor vectors of the $\underline{b}$ or varimax matrix, rather than getting one vector at a time. This method of solution should give exactly the same results, within limits of decimal error, as Kaiser's (1959) computational procedure. It appears to have the advantage that the computations are self-correcting and that it does not accumulate decimal error. The time required for the computations, as compared with the Kaiser method, has not been accurately determined, but it appears that for small matrices the Kaiser method may be slightly faster, and for larger matrices this method may be slightly faster.

One of the characteristics of the method is that each iteration requires the basic structure solution of the matrix whose order is equal to the number of factors. For high speed computers, however, this is not a serious restriction, since the number of factors would ordinarily not be over 10 or 15 at most, and available computer programs are extremely rapid for calculating the basic structure factors of matrices of this order.

18.4.2 Computational Equations

18.4.2a Definition of Notation

$\underline{a}$, $\underline{H}$, $\underline{b}$, $\underline{b^{(2)}}$, $\underline{b^{(3)}}$ are the same as in Section 18.3.2a.

18.4.2b The Equations

$$_1\beta = {}_1b^{(3)} - {}_1b \frac{D_{{}_1b'{}_1b}}{n} \tag{18.4.1}$$

$$_1C = a' \, {}_1\beta \tag{18.4.2}$$

$$_1C' \, {}_1C = {}_1Q \, {}_1\Delta^2 \, {}_1Q' \tag{18.4.3}$$

$$_2H = (({}_1C \, {}_1Q) \, {}_1\Delta^{-1}) \, {}_1Q' \tag{18.4.4}$$

$$_2b = a \, {}_2H \tag{18.4.5}$$

$$_i\beta = {}_ib^{(3)} - {}_ib \frac{D_{{}_ib'{}_ib}}{n} \tag{18.4.6}$$

$$_iC = a' \, {}_i\beta \tag{18.4.7}$$

$$_iC' \, {}_iC = {}_iQ \, {}_i\Delta^2 \, {}_iQ' \tag{18.4.8}$$

$$_{i+1}H = (({}_iC \, {}_iQ) \, {}_i\Delta^{-1}) \, {}_iQ' \tag{18.4.9}$$

$$_{i+1}b = a \, {}_{i+1}H \tag{18.4.10}$$

18.4.3 Computational Instructions. We begin with a first approximation to the simple structure factor loading matrix. This could be the normalized arbitrary factor matrix itself. It may be better to start with a more accurate approximation, such as some binary hypothesis method, as discussed in the previous chapter.

The first step in the computational cycles is indicated in Eq. (18.4.1). On the right hand side of this equation we have as the first term a matrix whose elements are the cubes of the elements of the first approximation to the $\underline{b}$ matrix. The second term on the right is obtained by postmultiplying the first approximation to the $\underline{b}$ matrix by a diagonal matrix. This diagonal matrix consists of the diagonal elements of the minor product moment of this approximation to the $\underline{b}$ matrix, divided by $\underline{n}$, the number of variables.

The next step is indicated in Eq. (18.4.2). Here we have the first approximation to a $\underline{C}$ matrix which is the minor product moment obtained by postmultiplying the transpose of the $\underline{a}$ factor loading matrix by the $\underline{\beta}$ matrix of Eq. (18.4.1).

The next set of computations is indicated by Eq. (18.4.3). We get the minor product moment of the matrix calculated in Eq. (18.4.2) and find its basic structure factors $\underline{Q}$ and $\underline{\Delta}^2$, as indicated on the right hand side of this equation.

We then get the second approximation to the transformation matrix, as indicated in Eq. (18.4.4). Here we postmultiply the $\underline{C}$ matrix of Eq. (18.4.2) successively by the factors $\underline{Q}$, $\underline{\Delta}^{-1}$, and $\underline{Q}'$. This second approximation to the $\underline{H}$ matrix is now orthonormal.

The second approximation to the $\underline{b}$ matrix is given in Eq. (18.4.5). This is the factor loading matrix $\underline{a}$ postmultiplied by the matrix of Eq. (18.4.4).

The general equations for the $\underline{i}$+1 approximations to the $\underline{H}$ and the $\underline{b}$ matrices are given in Eqs. (18.4.6) through (18.4.10), which are analogous to Eqs. (18.4.1) through (18.4.5). The subscripts 1 and 2 have been replaced by $\underline{i}$ and $\underline{i}$+1.

18.4.4 Numerical Example. We begin with the same factor loading matrix in this numerical example as in the previous one.

Table 18.4.1 gives the final approximation to the varimax transformation matrix $\underline{H}$. Intermediate approximations are not given, although they could be readily outputed from corresponding Fortran program.

Table 18.4.2 gives the final approximation to the varimax factor loading matrix. Here, too, the outputing of intermediate approximations may be readily inserted in the Fortran program. It can be seen that, aside from the order of the factors, the loadings do not differ markedly from those in the previous section. With other data the results may differ more for the two methods. As in the previous section, the matrix is normal by rows.

Table 18.4.1 - Final Approximation to Transformation Matrix $\underline{H}$ for Simultaneous Factor Varimax Solution

0.664	-0.472	-0.580
0.514	0.851	-0.105
0.545	-0.229	0.808

Table 18.4.2 - Final Approximation to Varimax Factor Matrix $\underline{b}$ for Simultaneous Factor Varimax Solution

0.980	0.001	0.197
0.972	-0.017	0.234
0.961	-0.223	0.164
-0.084	0.969	-0.233
-0.026	0.998	-0.049
-0.100	0.991	-0.085
0.168	-0.114	0.979
0.175	-0.008	0.984
0.318	-0.351	0.881

18.5 The Successive Factor General Varimax

18.5.1 Characteristics of the Method. This method is like the one discussed in Section 18.3, except that now instead of maximizing the variance of the squared factor vector elements, we solve for one vector at a time so as to maximize the variance of some power of the elements in which the power is the ratio of an even number to a smaller odd number. The computations are essentially the same as for Section 18.3 with the difference that, having chosen a particular power, we have the problem of finding the required powers of elements, either by means of tables of logarithmic and exponential functions or by means of computer program statements.

18.5.2 Computational Equations

18.5.2a Definition of Notation

${}_{\underline{1}}\underline{a}$, $\underline{b}$, and $\underline{H}$ are the same as in Section 18.3.2a.

18.5.2b The Equations

$$ {}_s\beta_{.1} = {}_sb_{.1}^{(2k-1)} - {}_sb_{.1}^{(k-1)} \frac{1'\,{}_sb_{.1}^{(k)}}{n} \tag{18.5.1} $$

$$ {}_sU_{.1} = {}_1a'\,{}_s\beta_{.1} \tag{18.5.2} $$

$$ {}_s\alpha_1 = \sqrt{{}_sU'_{.1}\,{}_sU_{.1}} \tag{18.5.3} $$

$$ {}_sH_{.1} = \frac{{}_sU_{.1}}{{}_s\alpha_1} \tag{18.5.4} $$

$$_{s+1}b_{.1} = {}_1a \; {}_sH_{.1} \tag{18.5.5}$$

$$_2a = {}_1a - b_{.1} H'_{.1} \tag{18.5.6}$$

$$_{i+1}a = {}_ia - b_{.i} H'_{.i} \tag{18.5.7}$$

18.5.3 Computational Instructions. In this model we may let the power of the transformed factor loading elements be any positive number $\underline{k}$ greater than unity which may be expressed as the ratio of an even to an odd integer.

We begin with Eq. (18.5.1). Here we indicate the $\underline{s}$ approximation to a $\underline{\beta}$ vector, which is analogous to the $\underline{\beta}$ vector which we calculated in Section 18.3. We assume that some approximation to the transformed vector is available. As a matter of fact, we can use the methods of Section 18.3 to get this approximation. We raise the elements to the $2\underline{k}-1$ power to get the first term on the right of Eq. (18.5.1), in which the subscript $\underline{s}$ takes the value 1. The second term on the right consists of the approximation vector with elements raised to the $\underline{k}-1$ power and multiplied by a scalar which is the mean of the $\underline{k}$th power elements of the vector.

Eq. (18.5.2) is obtained by postmultiplying the transpose of the arbitrary factor loading matrix by the $\underline{s}$ approximation to the $\underline{\beta_{.1}}$ vector to give an $\underline{s}$ approximation to a $\underline{U_{.1}}$ vector.

A scalar quantity is then calculated as in Eq. (18.5.3), which is the square root of the minor product moment of the vector calculated in Eq. (18.5.2).

The vector calculated in Eq. (18.5.2) is divided by the scalar calculated in Eq. (18.5.3) to give the $\underline{s}$ approximation to the first transformation vector of $\underline{H}$, as indicated in Eq. (18.5.4).

The $\underline{s+1}$ approximation to the first factor loading $\underline{b}$ vector is given by Eq. (18.5.5). This is the factor loading matrix postmultiplied by the vector calculated in Eq. (18.5.4). When the iterations stabilize sufficiently, as indicated by some tolerance limit set on the $\underline{\alpha}$ scalar of Eq. (18.5.3), we may take the resulting $\underline{b_{.1}}$ approximation as the first simple structure factor vector.

We then calculate a residual matrix as in Eq. (18.5.6). This is obtained by subtracting the major product of the first simple structure factor loading vector and the corresponding transformation vector $\underline{H}$ from the arbitrary factor matrix $\underline{a}$.

We now operate on this new matrix with Eqs. (18.5.1) through (18.5.5) in exactly the same way as we did on the original matrix. To get a first approximation we may use the method of Section 18.3.

The general equation for the $\underline{i}$+1 residual of the arbitrary factor loading matrix $\underline{a}$ is given by Eq. (18.5.7).

18.5.4 Numerical Example. We begin with the same data as in Section 18.3.4 and take $\underline{k} = \frac{10}{3}$. Here we indicate only the results of the method. The intermediate computations are not given.

Table 18.5.1 is the final approximation to the orthonormal transformation matrix.

Table 18.5.2 is the final approximation to the varimax matrix $\underline{b}$. It must be remembered that this is the transformation such that the variance of the $\frac{10}{3}$ power of the elements in the b matrix is a maximum for each column--with the restriction, of course, that the columns were obtained one at a time. It is of interest to compare this matrix with Table 18.3.3 of Section 18.3.4.

Table 18.5.1 - Final Approximation to Transformation Matrix $\underline{H}$ for Successive Factor General Varimax Method with $\underline{k} = \frac{10}{3}$

0.735	0.412	0.538
0.089	-0.846	0.526
-0.672	0.339	0.659

Table 18.5.2 - Final Approximation to Varimax Factor Matrix $\underline{b}$ for Successive Factor General Varimax Method with $\underline{k} = \frac{10}{3}$

0.358	-0.003	0.934
0.395	0.010	0.919
0.349	0.223	0.910
0.356	0.934	0.003
0.171	0.985	-0.026
0.218	0.975	0.042
1.000	-0.004	-0.001
0.994	-0.110	0.010
0.957	0.246	0.154

18.6 Simultaneous Factor General Varimax

18.6.1 Characteristics of the Method. This method is the same as the one discussed in Section 18.4, except that now we maximize the variance of any even power of the transformed elements we wish, as long as the power is greater than 1. The method, like that of Kaiser (1958), solves simultaneously for all of the transformed simple structure vectors, rather than for one at a time. It has the advantage that decimal error is not accumulated as it is in Kaiser's method. It has the disadvantage that, if the powers are not whole numbers, either tables or library programs for computers must be used in calculating these powers. However, this is not a serious disadvantage with the computers, since simple Fortran statements may be written for any specified power.

18.6.2 Computational Equations

18.6.2a Definition of Notation

$\underline{a}$, $\underline{b}$, and $\underline{H}$ are the same as in Section 18.3.2a.

$\underline{{}_s b}$ is the $\underline{s}$ approximation to $\underline{b}$.

$\underline{{}_s b^{(2k-1)}}$ is a matrix whose elements are those of $\underline{{}_s b}$ raised to the $2\underline{k}-1$ power.

$\underline{{}_s H}$ is the $\underline{s}$ approximation to the transformation matrix.

18.6.2b The Equations

$$ {}_s\beta = {}_sb^{(2k-1)} - {}_sb^{(k-1)} \frac{D_{1'{}_sb^{(k)}}}{n} \tag{18.6.1} $$

$$ {}_sC = a' \, {}_s\beta \tag{18.6.2} $$

$$ {}_sC' \, {}_sC = {}_sQ \, {}_s\Delta^2 \, {}_sQ' \tag{18.6.3} $$

$$ {}_sH = (({}_sC \, {}_sQ) \, {}_s\Delta^{-1}) \, {}_sQ' \tag{18.6.4} $$

$$ {}_{s+1}b = a \, {}_sH \tag{18.6.5} $$

18.6.3 Computational Instructions. In the example of the model discussed here, we shall again assume that it is the variance of the kth power of the elements in the transformed factor vectors which we wish to maximize.

We may begin with the arbitrary factor loading matrix itself, presumably normalized by rows. We consider Eq. (18.6.1) for the prescript s equal to 1. The first term on the right is a matrix whose elements are the $2k$-1 powers of the elements in ${}_1b$. The second term on the right has for the first factor a matrix whose elements are the k-1 powers of the elements in b. This matrix is postmultiplied by a diagonal matrix whose elements are the means of columns of a matrix obtained by raising the elements of b to the kth power.

Eq. (18.6.2) indicates the computation of a matrix C. This is the transpose of the arbitrary factor matrix, postmultiplied by the β matrix of Eq. (18.6.1).

Eq. (18.6.3) is the minor product moment of the matrix obtained in Eq. (18.6.2), and also indicates the solution for the basic structure factors,

$\underline{Q}$ and $\underline{\Delta}^2$, of this matrix. This involves, as in Section 18.4, basic structure computations outlined in earlier chapters.

The calculation of the $\underline{s}$ approximation to the $\underline{H}$ matrix is given in Eq. (18.6.4). This is obtained by postmultiplying in turn the $\underline{C}$ matrix of Eq. (18.6.2) by the $\underline{Q}$ matrix shown in Eq. (18.6.3), by the inverse of the $\underline{\Delta}$ matrix of Eq. (18.6.3), by the transpose of the $\underline{Q}$ matrix of Eq. (18.6.3).

The $\underline{s}+1$ approximation to the transformed factor loading matrix is given in Eq. (18.6.5). This is the arbitrary factor loading matrix $\underline{a}$ multiplied by the matrix of Eq. (18.6.4). A stabilization limit may be set on the sum of the elements of the basic diagonal in Eq. (18.6.3). These will in general increase asymptotically to an upper limit.

18.6.4 Numerical Example. We begin with the same principal axis matrix as in the preceding sections.

Table 18.6.1 gives the final approximation for the transformation matrix $\underline{H}$.

Table 18.6.2 gives the final approximation for the varimax factor matrix $\underline{b}$. The factors can readily be identified with those from previous solutions in this chapter, even though the results expectedly differ by more than decimal accuracy.

Table 18.6.1 - Final Approximation to Transformation Matrix H for the Simultaneous Factor General Varimax Solution with $k = \frac{10}{3}$

0.670	-0.476	-0.569
0.513	0.851	-0.108
0.536	-0.220	0.815

Table 18.6.2 - Final Approximation to the Varimax Factor Matrix b for the Simultaneous Factor General Varimax Solution with $k = \frac{10}{3}$

0.982	0.002	0.188
0.974	-0.017	0.224
0.963	-0.223	0.153
-0.089	0.971	-0.223
-0.028	0.999	-0.040
-0.104	0.992	-0.076
0.177	-0.123	0.976
0.185	-0.017	0.983
0.328	-0.358	0.874

18.7 Mathematical Proofs

18.7.1 Successive Factor Varimax

Let

$$b = a\,H \tag{18.7.1}$$

and $\underline{b^{(2)}}$ be a matrix whose elements are the squares of those in $\underline{b}$. Consider the function

$$b^{(2)\prime}_{.1}\left(I - \frac{1\,1'}{n}\right) b^{(2)}_{.1} = \phi_1 = \max \tag{18.7.2}$$

This is the well known varimax criterion of Kaiser (1958) which maximizes the variance of the squared factor loading vectors in $\underline{b}$. We impose the restriction that

$$H'_{.1}\,H_{.1} = 1 \tag{18.7.3}$$

We let

$$D_{b_{.1}}\,1 = b_{.1} \tag{18.7.4}$$

From Eqs. (18.7.1), (18.7.2), (18.7.3), and (18.7.4) we write

$$\psi_1 = H'_{.1}\,a'\,D_{b_{.1}}\left(I - \frac{1\,1'}{n}\right) D_{b_{.1}}\,a\,H_{.1} - H'_{.1}\,H_{.1}\,\lambda_1 \tag{18.7.5}$$

where $\underline{\lambda_1}$ is a Lagrangian scalar.

Differentiating Eq. (18.7.5) symbolically with respect to $\underline{H'_{.1}}$ and equating to 0, we have

$$\frac{\partial\,\psi_1}{\partial\,H'_{.1}} = 2\left[\left(a'\,D_{b_{.1}}\left(I - \frac{1\,1'}{n}\right) D_{b_{.1}}\,a\,H_{.1} - H_{.1}\,\lambda_1\right]\right. = 0 \tag{18.7.6}$$

From Eqs. (18.7.1) and (18.7.6)

$$a' \left(b^{(3)}_{.1} - b_{.1} \frac{b'_{.1}\, b_{.1}}{n} \right) = H_{.1}\, \lambda_1 \tag{18.7.8}$$

Let

$$\beta_{.1} = b^{(3)}_{.1} - b_{.1} \frac{b'_{.1}\, b_{.1}}{n} \tag{18.7.8}$$

From Eqs. (18.7.7) and (18.7.8)

$$a'\, \beta_{.1} = H_{.1}\, \lambda_1 \tag{18.7.9}$$

We start with a first approximation by considering, say,

$$W = \frac{a'\, 1}{\sqrt{1'\, a\, a'\, 1}} \tag{18.7.10}$$

$$_1V = a\, W \tag{18.7.11}$$

and find the smallest element in $\underline{_1V}$, say $\underline{_1V_L}$. We may then let

$$_0b_{.1} = a\, a_L. \tag{18.7.12}$$

We have as the general iteration equations

$$_s\beta_{.1} = {_s}b^{(3)}_{.1} - {_s}b_{.1} \frac{_s b'_{.1}\, {_s}b_{.1}}{n} \tag{18.7.13}$$

$$a'\, {_s}\beta_{.1} = {_s}U_{.1} \tag{18.7.14}$$

$$_sH_{.1} = \frac{_sU_{.1}}{\sqrt{_sU'_{.1}\, {_s}U_{.1}}} \tag{18.7.15}$$

$$_{s+1}b_{.1} = a\, {_s}H_{.1} \tag{18.7.16}$$

From Eqs. (18.7.1) and (18.7.6)

$$a' \left(b^{(3)}_{.1} - b_{.1} \frac{b'_{.1}\, b_{.1}}{n} \right) = H_{.1}\, \lambda_1 \tag{18.7.8}$$

Let

$$\beta_{.1} = b^{(3)}_{.1} - b_{.1} \frac{b'_{.1}\, b_{.1}}{n} \tag{18.7.8}$$

From Eqs. (18.7.7) and (18.7.8)

$$a'\, \beta_{.1} = H_{.1}\, \lambda_1 \tag{18.7.9}$$

We start with a first approximation by considering, say,

$$W = \frac{a'\, 1}{\sqrt{1'\, a\, a'\, 1}} \tag{18.7.10}$$

$${}_1V = a\, W \tag{18.7.11}$$

and find the smallest element in $\underline{{}_1V}$, say $\underline{{}_1V_L}$. We may then let

$${}_0b_{.1} = a\, a_L. \tag{18.7.12}$$

We have as the general iteration equations

$${}_s\beta_{.1} = {}_sb^{(3)}_{.1} - {}_sb_{.1} \frac{{}_sb'_{.1}\, {}_sb_{.1}}{n} \tag{18.7.13}$$

$$a'\, {}_s\beta_{.1} = {}_sU_{.1} \tag{18.7.14}$$

$${}_sH_{.1} = \frac{{}_sU_{.1}}{\sqrt{{}_sU'_{.1}\, {}_sU_{.1}}} \tag{18.7.15}$$

$${}_{s+1}b_{.1} = a\, {}_sH_{.1} \tag{18.7.16}$$

To get $\underline{H_{.2}}$ we require that

$$H'_{.2} H_{.1} = 0 \tag{18.7.17}$$

$$H'_{.2} H_{.2} = 1 \tag{18.7.18}$$

We could then write

$$\psi_2 = \phi_2 - 2 H'_{.2} H_{.1} \lambda_{12} - H'_{.2} H_{.2} \lambda_2 \tag{18.7.19}$$

where $\underline{\phi_2}$ is analogous to Eq. (18.7.2), and where $\underline{\lambda_{12}}$ and $\underline{\lambda_2}$ are Lagrangian multipliers.

Differentiating Eq. (18.7.19) symbolically with respect to $\underline{H'_{.2}}$ and equating to 0 gives, after solving for $\underline{\lambda_{12}}$ by means of Eq. (18.7.17),

$$(I - H_{.1} H'_{.1})\, a' \beta_{.2} = H_{.2} \lambda_2 \tag{18.7.20}$$

where

$$\beta_{.2} = b^{(3)}_{.2} - b_{.2} \frac{b'_{.2} b_{.2}}{n} \tag{18.7.21}$$

But from Eq. (18.7.1)

$$(I - H_{.1} H'_{.1})\, a' = a' - H_{.1} b'_{.1} \tag{18.7.22}$$

If we let

$${}_2a = a - b_{.1} H'_{.1} \tag{18.7.23}$$

we may write Eq. (18.7.20)

$${}_2a' \beta_{.2} = H_{.2} \lambda_2 \tag{18.7.24}$$

We may now solve for $\underline{H_{.2}}$ iteratively as we did for $\underline{H_{.1}}$, except that we use Eqs. (18.7.21) and (18.7.24). As a first approximation to $\underline{H_{.2}}$ we consider

$$ {}_{2}V = {}_{1}V + b_{.1} \tag{18.7.25}$$

Find the smallest value in $\underline{{}_{2}V}$, say, $\underline{{}_{s}V_{L_2}}$, and let

$$ {}_{0}H_{.2} = {}^{a}L_{2}. \tag{18.7.26}$$

In general, then,

$$ {}_{s}b_{.i} = {}_{i}a \; {}_{s}H_{.i} \tag{18.7.27}$$

$$ {}_{s}\beta_{.i} = {}_{s}b^{(3)}_{.i} - {}_{s}b_{.i} \; \frac{{}_{s}b'_{.i} \; {}_{s}b_{.i}}{n} \tag{18.7.28}$$

$$ {}_{i}a' \; {}_{s}\beta_{.i} = {}_{s}U_{.i} \tag{18.7.29}$$

$$ {}_{s}H_{.i} = \frac{{}_{s}U_{.i}}{\sqrt{{}_{s}U'_{.i} \; {}_{s}U_{.i}}} \tag{18.7.30}$$

$$ {}_{i+1}a = {}_{i}a - b_{.i} \; H'_{.i} \tag{18.7.31}$$

18.7.2 Simultaneous Factor Varimax

Let

$$ b = a\,H \tag{18.7.32}$$

and $\underline{b^{(2)}}$ be a matrix whose elements are the squares of those in $\underline{b}$. Consider the function

$$ b^{(2)\prime}_{.i} \left(I - \frac{1\,1'}{n}\right) b^{(2)}_{.i} = \phi_{i} = \max \tag{18.7.33}$$

We require that

$$H' H = I \qquad (18.7.34)$$

From Eqs. (18.7.32), (18.7.33), and (18.7.34) we write

$$\psi_i = H'_{.i} a' D_{b_{.i}} \left(I - \frac{1 1'}{n}\right) D_{b_{.i}} a H_{.i} - H'_{.i} H_{.i} \lambda_{ii} + \sum_{\substack{j=1 \\ i \neq j}}^{n} H'_{.i} H_{.j} \lambda_{ij} \qquad (18.7.35)$$

where the $\underline{\lambda_{ij}}$ are Lagrangian multipliers.

Let

$$\beta_{.i} = b^{(3)}_{.i} - b_{.i} \frac{b'_{.i} b_{.i}}{n} \qquad (18.7.36)$$

Differentiating Eq. (18.7.35) symbolically with respect to $\underline{H'_{.i}}$, equating to 0, and using Eq. (18.7.36), we have

$$a' \beta_{.i} - H \lambda_{.i} = 0 \qquad (18.7.37)$$

Or, setting up the complete matrix,

$$a' \beta - H \lambda = 0 \qquad (18.7.38)$$

where now $\underline{\lambda}$ is a matrix of Lagrangian multipliers, and where because of Eq. (18.7.35),

$$\lambda' = \lambda \qquad (18.7.39)$$

and from Eq. (18.7.36)

$$\beta = b^{(3)} - b \frac{D_{b'} b}{n} \qquad (18.7.40)$$

From Eq. (18.7.38)

$$a' \beta \lambda^{-1} = H \tag{18.7.41}$$

Let

$$a' \beta = P \Delta Q' \tag{18.7.42}$$

be the basic structure of $\underline{a' \beta}$.

From Eq. (18.7.42)

$$P \Delta Q' \lambda^{-1} = H \tag{18.7.43}$$

Because of Eqs. (18.7.34) and (18.7.39), the only $\underline{\lambda^{-1}}$ which will satisfy Eq. (18.7.43) is

$$\lambda^{-1} = Q \Delta^{-1} Q' \tag{18.7.44}$$

From Eqs. (18.7.43) and (18.7.44)

$$H = P Q' \tag{18.7.45}$$

We let

$$C = a' \beta \tag{18.7.46}$$

From Eqs. (18.7.42) and (18.7.46)

$$Q \Delta^2 Q' = C' C \tag{18.7.47}$$

From Eqs. (18.7.42) and (18.7.46)

$$P = C\,Q\,\Delta^{-1} \tag{18.7.48}$$

From Eqs. (18.7.45) and (18.7.48)

$$H = C\,Q\,\Delta^{-1}\,Q' \tag{18.7.49}$$

The iteration procedure is as follows. Given the $\underline{i}$th approximation to $\underline{b}$, then

$$_i\beta = {}_ib^{(3)} - {}_ib\,\frac{D_{{}_ib'\,{}_ib}}{n} \tag{18.7.50}$$

$$_iC = a'\,{}_i\beta \tag{18.7.51}$$

$$_iQ\,{}_i\Delta^2\,{}_iQ' = {}_iC'\,{}_iC \tag{18.7.52}$$

$$_{i+1}H = {}_iC\,{}_iQ\,{}_i\Delta^{-1}\,{}_iQ' \tag{18.7.53}$$

$$_{i+1}b = a\,{}_{i+1}H \tag{18.7.54}$$

18.7.3 Successive Factor General Varimax

Let

$$b_{.1} = a\,H_{.1} \tag{18.7.55}$$

$$k = \frac{2\,m_1}{2\,m_2 - 1} \tag{18.7.56}$$

where $\underline{m_1}$ and $\underline{m_2}$ are integers and $\underline{m_2} \lessgtr \underline{m_1}$. Let $\underline{b_{.1}^{(k)}}$ be the vector whose elements are the $\underline{k}$th power of those in $\underline{b_{.1}}$. Because of Eq. (18.7.56) all elements of $\underline{b_{.1}^{(k)}}$ are nonnegative.

Consider

$$b_{.1}^{(k)\prime}\left(I - \frac{1\,1'}{n}\right) b_{.1}^{(k)} = \phi_1 = \max \tag{18.7.57}$$

with the constraint

$$H'_{.1} H_{.1} = 1 \tag{18.7.58}$$

Let

$$\psi_1 = \phi_1 - k H'_{.1} H_{.1} \lambda_1 \tag{18.7.59}$$

where λ_1 is a Lagrangian scalar.

From Eqs. (18.7.55) and (18.7.57)

$$H'_{.1} a' D_{b_{.1}}^{k-1} (I - \frac{1 1'}{n}) D_{b_{.1}}^{k-1} a H_{.1} = \phi_1 \tag{18.7.60}$$

From Eqs. (18.7.59) and (18.7.60) we may write the maximizing equation

$$\frac{\partial \psi}{\partial H'_{.1}} = 2 k [a' D_{b_{.1}}^{k-1} (I - \frac{1 1'}{n}) D_{b_{.1}}^{k-1} a H_{.1} - H_{.1} \lambda_1] = 0 \tag{18.7.61}$$

We let

$$\beta_{.1} = b_{.1}^{(2k-1)} - b_{.1}^{(k-1)} \frac{1' b_{.1}^{(k)}}{n} \tag{18.7.62}$$

From Eqs. (18.7.61) and (18.7.62)

$$a' \beta_{.1} = H_{.1} \lambda_1 \tag{18.7.63}$$

We may begin with some arbitrary $b_{.1}$, say, ${}_1b_{.1}$, and set up the iterative equations

$${}_s\beta_{.1} = {}_sb_{.1}^{(2k-1)} - {}_sb_{.1}^{(k-1)} \frac{1' {}_sb_{.1}^{(k)}}{n} \tag{18.7.64}$$

$$a' {}_s\beta_{.1} = {}_sU_{.1} \tag{18.7.65}$$

$$ {}_s\alpha_1 = \sqrt{{}_sU'_{.1}\ {}_sU_{.1}} \qquad (18.7.66) $$

$$ {}_sH_{.1} = \frac{{}_sU_{.1}}{{}_s\alpha_1} \qquad (18.7.67) $$

$$ {}_sb_{.1} = a\ {}_sH_{.1} \qquad (18.7.68) $$

after α_1 stabilizes, we calculate

$$ {}_2a = a - b_{.1}\ H'_{.1} \qquad (18.7.69) $$

This is then substituted for a in Eqs. (18.7.65) and (18.7.68) to get $b_{.2}$. The general equation for Eq. (18.7.69) is, of course,

$$ {}_{i+1}a = {}_ia - b_{.i}\ H'_{.i} \qquad (18.7.70) $$

To calculate the elements in ${}_sb_{.i}$ raised to the respective powers, we proceed as follows. We let B be any element of ${}_sb_{.i}$. Now because of Eq. (18.7.56)

$$ k - 1 = \frac{2\ (m_1 - m_2) + 1}{2\ m_2 - 1} \qquad (18.7.71) $$

and

$$ 2\ k - 1 = \frac{2\ (2\ m_1 - m_2) + 1}{2\ m_2 - 1} \qquad (18.7.72) $$

From Eq. (18.7.71) all B^{k-1} and B^{2k-1} have the same sign as B. From Eq. (18.7.56) all B^k are positive. Hence to calculate the several required powers of B we have

$$ B^k = \text{antilog}\ [k \log |B|] \qquad (18.7.73) $$

$$B^{k-1} = \text{antilog}\,[(k - 1)\log |B|]\,.\frac{|B|}{B} \qquad (18.7.74)$$

$$B^{2k-1} = \text{antilog}\,[(2\,k - 1)\log |B|]\,\frac{|B|}{B} \qquad (18.7.75)$$

These three powers of $\underline{B}$ may be readily calculated from tables or from standard library programs for computer installations.

18.7.4 Simultaneous Factor General Varimax

Let

$$b = a\,H \qquad (18.7.76)$$

$$k = \frac{2\,m_1}{2\,m_2 - 1} \qquad (18.7.77)$$

where $\underline{m_1}$ and $\underline{m_2}$ are integers and $\underline{m_2} \lesseqgtr \underline{m_1}$. Consider, then, the generalization of Kaiser's criterion

$$tr\left[b^{(k)\prime}\,(I - \frac{1\,1'}{n})\,b^{(k)}\right] = \max = \phi \qquad (18.7.78)$$

with the constraint

$$H'\,H = I \qquad (18.7.79)$$

From Eqs. (18.7.76), (18.7.77), and (18.7.78)

$$\Sigma\,H'_{.i}\,a'\,D^{k-1}_{b_{.i}}\,(I - \frac{1\,1'}{n})\,D^{k-1}_{b_{.i}}\,a\,H_{.i} = \phi \qquad (18.7.80)$$

Consider

$$\psi = \phi - k\,tr\,(H'\,H\,\lambda) \qquad (18.7.81)$$

where $\underline{\lambda}$ is a symmetric matrix of Lagrangian multipliers.

From Eqs. (18.7.80) and (18.7.81) we may write

$$\frac{\partial \psi}{\partial H'_{.i}} = 2k \left[a' D_{b_{.i}}^{k-1} \left(I - \frac{1\,1'}{n}\right) D_{b_{.i}}^{k-1} a H_{.i} - H \lambda_{.i} \right] = 0 \qquad (18.7.82)$$

Let

$$\beta_{.i} = b_{.i}^{(2k-1)} - b_{.i}^{(k-1)} \frac{1' b_{.i}^{(k)}}{n} \qquad (18.7.83)$$

From Eqs. (18.7.82) and (18.7.83)

$$a' \beta_{.i} - H \lambda_{.i} = 0 \qquad (18.7.84)$$

or for the matrix form

$$a' \beta = H \lambda \qquad (18.7.85)$$

By the methods of the previous section, we may let the basic structure of $\underline{a' \beta}$ be

$$a' \beta = P \Delta Q' \qquad (18.7.86)$$

From Eq. (18.7.85)

$$a' \beta \lambda^{-1} = H \qquad (18.7.87)$$

To satisfy Eqs. (18.7.87) and (18.7.79)

$$H = a' \beta Q \Delta^{-1} Q' \qquad (18.7.88)$$

We may begin with some arbitrary approximation to $\underline{b}$ in Eq. (18.7.76), say, $\underline{{}_1b}$. From $\underline{{}_1b}$ and Eq. (18.7.83) we calculate $\underline{{}_1\beta}$. Then from Eq. (18.7.86)

$$a' \; {}_1\beta = {}_1P \; {}_1\Delta \; {}_1Q' \tag{18.7.89}$$

and from Eq. (18.7.88)

$$ {}_1H = {}_1P \; {}_1Q' \tag{18.7.90}$$

In general, then,

$$ {}_s\beta_{.i} = {}_sb_{.i}^{(2k-1)} - {}_sb_{.i}^{(k-1)} \; \frac{1' \; {}_sb_{.i}^{(k)}}{n} \tag{18.7.91}$$

$$a' \; {}_s\beta = {}_sP \; {}_s\Delta \; {}_sQ' \tag{18.7.92}$$

$$ {}_sH = a' \; {}_s\beta \; {}_sQ \; {}_s\Delta^{-1} \; {}_sQ' \tag{18.7.93}$$

$$ {}_{s+1}b = a \; {}_sH \tag{18.7.94}$$

The several required powers of the $\underline{{}_sb_{.i}}$ in Eq. (18.7.91) can be calculated as in the previous section.

CHAPTER 19

DIRECT VARIMAX SOLUTIONS

We saw in Chapter 18 that we may begin with any arbitrary factor loading matrix, and transform it to a simple structure varimax factor loading matrix by means of a square orthonormal transformation. In this chapter, we shall see how we may avoid the intermediate step of first calculating an arbitrary factor loading matrix, such as in the principal axis, multiple group, group centroid, and variations of these methods.

19.1 Characteristics of the Method

19.1.1 No Arbitrary Factor Matrices Required. In the procedures outlined in the following sections, we shall see how we may operate directly on the correlation matrices, or the score matrices from which they are derived. Strangely enough, this seems to be a novel approach for factor analysts. The tradition has been to calculate first some arbitrary factor loading matrix, such as the principal axis, centroid, or other type of arbitrary factor loading matrix, whose major product moment gives a reasonable approximation to the correlation matrix. These arbitrary matrices are then transformed by the procedures outlined in the previous chapter or by other analytical, semianalytical, or graphical methods. Actually, the transition from operations upon the arbitrary matrices to achieve simple structure matrices, to that of a direct solution for the simple structure matrices from the correlation or data matrices, is perfectly natural, both from a logical and a mathematical point of view.

19.1.2 Analytical Methods. As a matter of fact, however, the methods of direct solution for simple structure factor matrices are applicable primarily to those procedures and rationales which use analytical, rather than graphical or judgmental, methods. It should be pointed out that the method for achieving simple structure by means of a binary hypothesis matrix, such as discussed in Chapter 17 for the special case of the multiple group method, might be regarded as a special case of a direct solution from the correlation matrix. Here, however, the rationale and procedure is essentially different from that of the methods to be discussed in this chapter.

All of the methods discussed in this chapter are based on the analytical procedures of the varimax method, and constitute applications of the rationale of this method to the correlation and data matrices. The direct solutions for simple structure matrices need not be limited to the varimax type of solution. However, as in the previous chapter, they will be so restricted in this chapter because of the practical difficulties encountered with other alternatives to the varimax solutions which have been developed and experimented with so far.

19.1.3 Rank Reduction Solutions. All of the solutions considered in this chapter are of a rank reduction type. This means that the rank of the residual matrix, following the solution of any factor vector, is one less than that of the previous residual matrix, or, in the case of a factor matrix, its rank is equal to that of the correlation matrix less the number of factors in the factor matrix.

19.1.4 Iterative Solutions. All of the solutions outlined below are of an iterative type. In this respect, they differ from the direct solution considered for the multiple group method in Chapter 17, in which a binary hypothesis was used. Here, we recall, no successive approximations were required, except for the orthonormal transformation.

All of the iterative solutions required for the various types of direct solutions outlined below can be shown to be special, though rather complicated, eigenvector or basic structure type solutions in which we have a symmetric matrix, some of whose elements are functions of its own basic orthonormal and basic diagonal elements. Therefore, the iterative procedure may be somewhat more involved than in the case of a straightforward basic structure solution, in which the elements of the symmetric matrix whose basic structure is desired are constant values. As a matter of fact, this type of basic structure solution is characteristic not only of the methods of this chapter, but also of those of the previous chapter, in which the solutions are not applied directly to the correlation or data matrices but to some arbitrary factor matrix.

19.1.5 Utilization of Information. One of the distinguishing characteristics of the methods outlined in this chapter is that, in a sense, more of the information in the correlation or data matrix is utilized than in the methods of Chapter 18. Implicit in the methods of both Chapters 17 and 18 is the assumption that the factor loading matrix accounts for all of the significant or systematic nonrandom information inherent in the correlation or data matrices. This assumption may not be valid in many cases.

In the methods outlined below, all of the information in the data or correlation matrices is utilized in the determination of the simple structure varimax factor matrices or vectors. This may be regarded as an advantage, from the point of view of information utilization; or it may be regarded as a disadvantage, if one takes the position that the information left over after the major product moment of the factor loading matrix is subtracted from the correlation matrix essentially represents error or random variation. In this latter case, one may argue that the application of direct varimax methods to the correlation or the data matrices may be spuriously affected by such random variation. A great deal more theoretical and experimental work needs to be done before the relative validity of these alternative points of view can be established. In general, the criterion of invariance from one sample of entities and attributes to another would be a relevant consideration here. Presumably, if the direct methods turn out to yield more consistent results from one sample of entities to another, and from one sample of attributes to another, their superiority would be definitely indicated.

19.1.6 Results Different from Transformation Solutions. It must become obvious that one cannot expect exactly the same results from the direct methods outlined below as from the transformation solutions indicated in Chapters 17 and 18. Even for the direct methods of Chapter 18, in which we use the same varimax criteria and precisely the same models, one cannot expect to get exactly the same results as when the criteria are applied to the correlation or data matrices. The reason for this is, of course, that we utilize not only the information provided by some arbitrary factor loading matrix, but also information inherent in the data or correlation matrix which has not been reflected in the solution for the arbitrary factor loading matrix.

Only extensive research can tell which of the methods is better from the point of view of factorial invariance. There is, of course, the question of which factor loadings make more sense from the point of view of the particular discipline concerned. But as we have suggested earlier, the determination of whether or not the results make sense for a particular discipline is subjective. Until the concept is more objectively defined than it has been in the past, we cannot use the criterion of how much sense the results make as a basis for comparison of any of the methods of factor analysis.

19.2 Kinds of Methods

For convenience, we may group the various methods into four classes. These are (1) the successive factor varimax solution from the correlation matrix, (2) the simultaneous factor varimax solution from the correlation matrix, (3) the successive factor varimax solution from the data matrix, and (4) the simultaneous factor varimax solution from the data matrix. Each of these classes of solutions may, in the conventional manner, maximize the variance of the squared elements, or, more generally, the variance of some other even power, just as in Chapter 18. We shall consider first the conventional type solutions, and then the general type.

19.2.1 Solutions from the Correlation Matrix. As indicated in the previous section, the direct solutions may proceed either by operations on the correlation matrix or by operations directly on the data matrix. The methods based on operations on the correlation matrix may again be of two kinds, analogous to the two types of varimax rotations for arbitrary factor matrices. One of these is the successive factor vector method which obtains a single factor vector at a time. The other is the simultaneous factor matrix method, which iterates successively to the entire factor loading matrix for the particular number of factors hypothesized to be significant for a set of variables.

19.2.2 Solutions from the Data Matrix. One may bypass the calculation of the correlation matrix and operate directly upon the data matrix. Solutions of this class obviously cannot achieve a net saving in computations over the number required for the correlation matrix, since, as one may guess, the computations for the direct varimax from the data matrix involve more computations than solutions based on the correlation matrix itself. The question as to which of the methods is most economical from a computer or cost point of view depends on a number of factors. In general, if the number of cases is not vastly greater than the number of variables, one may save some time operating directly upon the data matrix.

In any case, the methods outlined here assume that the data matrix has been scaled so that the variables have means of 0 and variances of unity. It is possible, of course, to work out computational procedures so that the data matrix need not first be processed to yield a standardized metric. Such a computational procedure would incorporate the vector of means and the diagonal of standard deviations or variances. These methods, like those in Section 19.2.1, include the successive factor varimax model and the simultaneous factor matrix model, but here they are applied directly to the scaled data matrix.

19.2.3 The Successive Factor General Varimax Method. The conventional or Kaiser varimax method, as we know, maximizes the variance of the squared factor loadings. This rationale, we recall, may be generalized so that the variance of any even power function of the factor loading variable may be maximized. As indicated in Chapter 18, the power should be greater than unity. Otherwise, we may get into difficulty with reciprocals of very small values. The successive factor general varimax procedures may again be divided into two models. One of these operates on the correlation matrix, and the other operates directly on the data matrix.

19.2.4 Simultaneous Factor General Varimax. Just as we have the simultaneous general factor varimax method operating on the arbitrary factor loading matrix, so also we can apply this model directly to either the correlation matrix or the data matrix. In this chapter we shall consider its application to both the correlation matrix and the data matrix. The data matrix is assumed to be scaled so as to give means of 0 and variances of unity.

19.3 Successive Varimax Factor from Correlation Matrix

19.3.1 Characteristics of the Method. We are already familiar with some of the characteristics of the successive factor varimax method applied directly to the correlation matrix. It may be of interest to compare this method with others, such as the principal axis or centroid, with respect to the amount of variance accounted for by the successive factors. In the latter methods, we recall that the amount of variance accounted for tends to decrease, in general, with the successive factors calculated.

We cannot, however, assert that each varimax factor calculated from the correlation matrix does account for more of the variance than the subsequent one. We cannot even guarantee that the criterion of maximum variance of the squared factor loadings will be greater for a given factor vector than for one calculated subsequently. The order depends very largely on the characteristics of the matrix, and also on what is used as a first approximation for any particular factor loading vector.

A further reason for the uncertainty of the order of the factors, with respect to the amount of variance accounted for and the variance of their squared factor loadings, is the fact that we are dealing with an eigenvector or basic structure problem of a very complicated nature. As we have seen, the model involves symmetric matrices whose elements are functions of its eigenvectors and eigenvalues. We do not yet have available an adequate mathematical substructure for a satisfactory understanding of what determines the order in which the factors will appear.

19.3.2 Computational Equations

19.3.2a Definition of Notation

$_{i}R$ is the ith residual correlation matrix where $_{1}R$ is the correlation matrix itself.

$_{s}b_{.i}$ is the s approximation to the ith varimax factor vector.

$_{s}\gamma_{i}$ is the amount of variance accounted for by the s approximation to the ith varimax factor vector.

$_{s}\alpha_{i}$ is the varimax criterion for the s approximation to the ith varimax factor vector.

$_{s}b_{.i}^{(3)}$ is a factor vector whose elements are the cubes of those in $_{s}b_{.i}$.

19.3.2b The Equations

$$_{i}R = {}_{i-1}R - b_{.(i-1)}\, b'_{.(i-1)} \tag{19.3.1}$$

$$ {}_s\gamma_i = {}_s b'_{.i} \; {}_s b_{.i} \qquad (19.3.2) $$

$$ {}_s\beta_{.i} = {}_s b^{(3)}_{.i} - {}_s b_i \frac{{}_s\gamma_i}{n} \qquad (19.3.3) $$

$$ {}_s W_{.i} = {}_i R \; {}_s\beta_{.i} \qquad (19.3.4) $$

$$ {}_s\alpha_i = \sqrt{{}_s W'_{.i} \; {}_s\beta_{.i}} \qquad (19.3.5) $$

$$ {}_{s+1} b_{.i} = \frac{{}_s W_{.i}}{{}_s\alpha_i} \qquad (19.3.6) $$

$$ r = {}_1 R - I \qquad (19.3.7) $$

$$ \rho = r^{(2)} \qquad (19.3.8) $$

$$ U_{.0} = \rho 1 - \frac{(r\,1)^{(2)}}{n - 1} \qquad (19.3.9) $$

$U_{L\,0}$ is largest element in U $\qquad$ (19.3.10)

$$ {}_1 b_{.1} = {}_1 R_{.L} \qquad (19.3.11) $$

19.3.3 Computational Instructions. We begin with a correlation matrix. The meaning of Eq. (19.3.1) is as follows. If $\underline{i} = 1$, we have simply the correlation matrix. The right hand side of this equation will then be ignored, because the $\underline{i} - 1$ would be 0 and have no meaning, that is, it would not be defined. It is only for $\underline{i}$ greater than 1 that Eq. (19.3.1) has meaning.

We assume now that we have a first approximation to a factor loading vector which will be described in detail in Eqs. (19.3.7) through (19.3.11). Eqs. (19.3.2) through (19.3.6) describe the successive cycles for the s approximation to the ith factor loading vector.

Having given any approximation to the ith factor loading vector, its minor product moment is calculated as indicated in Eq. (19.3.2).

Next we calculate a vector as indicated in Eq. (19.3.3). This is the β vector with which we are already familiar. It is obtained by cubing the elements of the b vector, and subtracting from it the b vector multiplied by the scalar in Eq. (19.3.2) divided by n.

The next step is indicated by Eq. (19.3.4), which again is of course a general equation for the s approximation. It is the product of the ith residual correlation matrix postmultiplied by the β vector calculated in Eq. (19.3.3).

Next we calculate the scalar indicated by Eq. (19.3.5). This is the square root of the minor product moment of the vectors of Eqs. (19.3.3) and (19.3.4).

The s+1 approximation to the varimax factor loading vector for the ith factor is calculated in Eq. (19.3.6). This is the vector of Eq. (19.3.4) divided by the scalar of Eq. (19.3.5).

As yet there appears to be no completely satisfactory method for choosing a first approximation to any particular varimax factor loading vector, including the first one. However, the following method is recommended and should give good results in most cases.

Consider the correlation matrix with 0's in the diagonal, as indicated in Eq. (19.3.7).

First we square each element of the matrix $\underline{r}$ of Eq. (19.3.7), as indicated in Eq. (19.3.8).

We then calculate a $\underline{U}$ vector as indicated in Eq. (19.3.9). This is obtained by calculating a vector of the sums of the rows of the matrix in Eq. (19.3.8). From this is subtracted a vector of the squares of the sums of rows of the matrix in Eq. (19.3.7) divided by $\underline{n} - 1$, the number of variables. The $\underline{U}$ vector is therefore $\underline{n} - 1$ times the vector of the variances of the columns of the correlation matrix with the diagonal elements excluded.

In Eq. (19.3.10), we find the largest element in the $\underline{U}$ vector calculated in Eq. (19.3.9), and call this the $\underline{L}$ position.

We then take as the first approximation to the first varimax vector the $\underline{L}$th column of the correlation matrix, as indicated in Eq. (19.3.11).

Using Eqs. (19.3.2) through (19.3.6), we then calculate the first varimax factor loading vector by successive iterations until the $\underline{\alpha}$ scalar of Eq. (19.3.5) stabilizes to some prespecified degree of accuracy.

Then we calculate the first residual matrix by substituting 2 for the $\underline{i}$ subscript in Eq. (19.3.1). This residual matrix is simply the original correlation matrix, less the major product moment of the final approximation to the $\underline{b_{.1}}$ vector.

We now require a first approximation to the second varimax factor vector. We apply the procedures of Eqs. (19.3.7) through (19.3.11) to the residual matrix.

We continue with Eqs. (19.3.1) through (19.3.6), and generalizations of Eqs. (19.3.7) through (19.3.11), until enough factors have been extracted.

19.3.4 Numerical Example. In this and subsequent numerical examples in this chapter, the presentation of the results will not conform to that of the presentation of the methods themselves. Rather it will conform to a computer program sequence which is more efficient for performing the computations included in all of the methods of this chapter. Each numerical example section will include both the conventional case in which the variance of squared elements of the varimax factor vectors are maximized, and a special example of the general case for the $\frac{10}{3}$ power.

The same correlation matrix used in previous chapters will be used for all of those methods beginning with the correlation matrix. This is repeated for convenient reference in Table 19.3.1.

The same data matrix will be used for all methods operating directly on the data matrix. This matrix yields the correlation matrix of Table 19.3.1, so that the results obtained from the data matrix may be compared with those obtained from the correlation matrix.

Table 19.3.2 gives the data matrix.

Table 19.3.3 gives the varimax factor matrix for three factors obtained by the successive factor matrix method directly from the correlation matrix, for the conventional case in which the variances of the elements of the squared factor loading vectors are maximized.

Table 19.3.4 gives the varimax factor loading matrix for the successive factor general matrix, where the $\frac{10}{3}$ power, rather than the square of the elements, is used.

Table 19.3.1 - Correlation Matrix

1.000	0.829	0.768	0.108	0.033	0.108	0.298	0.309	0.351
0.829	1.000	0.775	0.115	0.061	0.125	0.323	0.347	0.369
0.768	0.775	1.000	0.272	0.205	0.238	0.296	0.271	0.385
0.108	0.115	0.272	1.000	0.636	0.626	0.249	0.183	0.369
0.033	0.061	0.205	0.636	1.000	0.709	0.138	0.091	0.254
0.108	0.125	0.238	0.626	0.709	1.000	0.190	0.103	0.291
0.298	0.323	0.296	0.249	0.138	0.190	1.000	0.654	0.527
0.309	0.347	0.271	0.183	0.091	0.103	0.654	1.000	0.541
0.351	0.369	0.385	0.369	0.254	0.291	0.527	0.541	1.000

Table 19.3.2 - Normalized Deviation Data Matrix

0.128	0.181	0.421	0.506	0.857	0.746	0.280	0.178	0.246
0.764	0.740	0.563	-0.387	-0.293	-0.202	0.261	0.281	0.043
-0.030	-0.046	0.014	0.147	-0.109	-0.135	0.640	0.682	0.661
-0.280	-0.351	-0.326	-0.023	-0.109	-0.186	0.083	0.091	-0.654
-0.336	-0.306	-0.429	-0.542	-0.006	-0.153	-0.428	0.056	-0.124
-0.276	-0.324	-0.271	-0.370	-0.225	0.035	0.129	-0.446	-0.124
0.057	-0.070	-0.016	0.006	0.152	-0.441	-0.166	-0.354	-0.033
-0.010	-0.140	0.326	-0.004	-0.258	-0.091	-0.410	-0.200	-0.101
-0.303	0.227	-0.014	0.029	-0.102	-0.125	-0.086	-0.106	-0.074
0.086	0.057	-0.003	0.161	0.073	0.134	-0.082	-0.067	-0.009
0.164	0.106	-0.124	0.234	-0.002	0.227	-0.072	0.025	0.156
0.036	-0.074	-0.142	0.242	0.023	0.192	-0.148	-0.119	0.011

Table 19.3.3 - Successive Factor Varimax Matrix from Correlation Matrix for $\underline{k} = 2$

0.937	0.003	0.092
0.933	0.023	0.127
0.887	0.175	0.063
0.098	0.782	0.130
0.042	0.921	0.026
0.100	0.894	0.044
0.234	0.129	0.837
0.243	0.058	0.903
0.300	0.247	0.527

Table 19.3.4 - Successive Factor General Varimax Matrix from Correlation Matrix for $\underline{k} = \frac{10}{3}$

0.947	-0.040	0.018
0.946	-0.013	0.057
0.880	0.140	-0.012
0.143	0.663	0.099
0.078	0.991	0.000
0.143	0.769	0.008
0.302	0.123	0.596
0.311	0.067	0.948
0.364	0.235	0.430

19.4 Simultaneous Varimax Matrix from the Correlation Matrix

19.4.1 Characteristics of the Method. This method is like the preceding one in that we operate directly on the correlation matrix rather than on the arbitrary factor matrix. Here, however, we iterate simultaneously to all of the factor loading vectors which we wish to solve for. It differs also from the previous method in that we have a less objective way for getting a first approximation to the factor loading vectors than we did in that method. It is also essentially different computationally from the previous method in that each iteration involves the solution for the basic structure of a symmetric matrix. The $\underline{Q}$ orthonormal and $\underline{\Delta}^2$ basic diagonal of this matrix are required in the successive approximations to the factor loading vectors. In this respect the method is analogous to the simultaneous method of Chapter 18.

19.4.2 Computational Equations

19.4.2a Definition of Notation

$\underline{R}$ is the correlation matrix.

$\underline{{}_s b}$ is the $\underline{s}$ approximation to the varimax factor matrix.

$\underline{{}_s b^{(3)}}$ is a matrix whose elements are the cubes of those in $\underline{{}_s b}$.

$\underline{D_{{}_s b' {}_s b}}$ is a diagonal matrix whose elements are the diagonals of the minor product moment of $\underline{{}_s b}$.

$\underline{{}_s Q}$ is a basic orthonormal.

$\underline{{}_s \Delta^2}$ is a basic diagonal.

19.4.2b The Equations

$$ {}_s\beta = {}_sb^{(3)} - {}_sb \frac{D_{{}_sb'\,{}_sb}}{n} \qquad (19.4.1) $$

$$ {}_sM = R\ {}_s\beta \qquad (19.4.2) $$

$$ {}_sG = {}_s\beta'\ {}_sM \qquad (19.4.3) $$

$$ {}_sQ\ {}_s\Delta^2\ {}_sQ' = {}_sG \qquad (19.4.4) $$

$$ {}_{s+1}b = (({}_sM\ {}_sQ)\ {}_s\Delta^{-1})\ {}_sQ' \qquad (19.4.5) $$

19.4.3 Computational Instructions. In this method we start with some arbitrary approximation to the varimax factor loading matrix. This may be simply the first $\underline{m}$ columns of the correlation matrix, where $\underline{m}$ is the number of factors. In any case, one must make an assumption as to the number of significant factors in the data matrix. It is better to overestimate than to underestimate the number, as some of the factors can later be rejected from the final stabilized varimax factor loading matrix if they do not seem interpretable or of sufficient importance.

If one has an hypothesis as to which of the variables represents which factor, he may select a variable to represent each factor. Then the columns of the correlation matrix corresponding to these variables will constitute the vectors of the first approximation to the varimax factor matrix.

We begin then with Eq. (19.4.1) in which the subscript $\underline{s}$ is taken as 1. This equation is similar to those we are already familiar with. The right hand side has for the first term a matrix whose elements are the cubes of the elements of the corresponding approximation to the factor loading matrix. We subtract from this the factor loading matrix itself, after scaling it by a diagonal matrix. This diagonal matrix is made up of the diagonals of the minor product moment of the current approximation to the factor loading matrix, and then divided by n, the number of variables.

Next we calculate the product indicated in Eq. (19.4.2). This is the correlation matrix postmultiplied by the matrix calculated in Eq. (19.4.1).

We then calculate the minor product of the matrices calculated in Eqs. (19.4.1) and (19.4.2). This is the matrix G indicated in Eq. (19.4.3).

Next we calculate the basic structure factors of the matrix G. This is indicated by Eq. (19.4.4).

Finally, for each approximation, we calculate the product of the four factors as indicated in Eq. (19.4.5). This is the product of the matrix of Eq. (19.4.2) postmultiplied first by the right orthonormal of the matrix in Eq. (19.4.4), then by the inverse of the square root of the basic diagonal of the matrix in Eq. (19.4.4), and finally by the left orthonormal of the matrix in Eq. (19.4.4), which is, of course, the transpose of the right orthonormal. This is true because G is a product moment matrix. We now have the s+1 approximation to the matrix of varimax factor loadings. These iterations continue until the approximation is sufficiently close.

A good criterion of convergence is the trace of the G matrix given by Eq. (19.4.3). Another criterion may be the trace of the minor product moment of the current approximation to the factor loading matrix. This would be the diagonal matrix in the right term of the right side of Eq. (19.4.1). This trace is simply the total amount of variance accounted for by any particular

$\underline{s}$ approximation to the factor loading matrix. As indicated earlier in this chapter, any approximation is a rank reduction solution, and therefore the larger this trace the greater the amount of variance accounted for.

19.4.4 Numerical Example. This example begins with the same correlation matrix given in Table 19.3.1.

Table 19.4.1 gives the first three varimax factor vectors for the simultaneous factor method applied to the correlation matrix for the conventional case of the squared elements.

Table 19.4.2 gives the first three varimax factor vectors for the simultaneous factor general method applied to the correlation matrix for the case of the $\frac{10}{3}$ power of the varimax factor elements. The results are not markedly different from those in Table 19.4.1.

Table 19.4.1 - Simultaneous Factor Varimax Matrix from Correlation Matrix for $\underline{k} = 2$

0.929	0.007	0.153
0.922	0.025	0.187
0.882	0.179	0.125
0.085	0.781	0.153
0.033	0.917	0.048
0.090	0.896	0.072
0.177	0.112	0.884
0.186	0.040	0.891
0.266	0.238	0.551

Table 19.4.2 - Simultaneous Factor General Varimax Matrix from Correlation Matrix for $\underline{k} = \frac{10}{3}$

0.944	0.018	0.121
0.935	0.024	0.148
0.857	0.194	0.112
0.080	0.989	0.116
0.037	0.670	0.049
0.100	0.645	0.092
0.189	0.123	0.974
0.242	0.088	0.633
0.306	0.292	0.442

19.5 Successive Factor Vector from Data Matrix

19.5.1 Characteristics of the Method. This method differs essentially from the preceding methods in that the calculation of the correlation matrix is not required. The computations proceed directly upon the data matrix. It is assumed that they have been previously scaled. This assumption is not imperative, however, since the computations could be modified to operate on a raw data matrix. The method differs also from the successive factor matrix method of Section 19.3 in that it is difficult to select a first approximation to a factor loading vector by objective means. Perhaps the simplest way to get the first approximation to the first factor loading vector is to assume that the first variable is not an extremely poor representation of the first varimax factor.

19.5.2 Computational Equations

19.5.2a Definition of Notation

$\underline{{}_i Z}$ is the $\underline{i}$th residual data matrix where $\underline{{}_1 Z}$ is the scaled data matrix with 0 means and unit variances.

$\underline{{}_s\gamma_i}$, $\underline{{}_s b_{.i}}$, $\underline{{}_s b^{(3)}_{.i}}$, and $\underline{{}_s\alpha_i}$ are the same as in Section 19.3.2a.

$\underline{{}_s Y_{.i}}$ is the $\underline{s}$ approximation to the $\underline{i}$th varimax factor score vector.

19.5.2b The Equations

$$ {}_i Z = {}_{i-1} Z - Y_{.(i-1)}\, b'_{.(i-1)} \qquad (19.5.1)$$

$$ {}_1 b_{.i} = {}_i Z'\, {}_i Z_{.} L_i \qquad (19.5.2)$$

$$ {}_s\gamma_i = {}_sb'_{.i}\ {}_sb_{.i} \tag{19.5.3} $$

$$ {}_s\beta_{.i} = {}_sb^{(3)}_{.i} - {}_sb_{.i}\ \frac{{}_s\gamma_i}{n} \tag{19.5.4} $$

$$ {}_sU_{.i} = {}_iZ\ {}_s\beta_{.i} \tag{19.5.5} $$

$$ {}_s\alpha_i = \sqrt{{}_sU'_{.i}\ {}_sU_{.i}} \tag{19.5.6} $$

$$ {}_sY_{.i} = \frac{{}_sU_{.i}}{{}_s\alpha_i} \tag{19.5.7} $$

$$ {}_{s+1}b_{.i} = {}_iZ'\ {}_sY_{.i} \tag{19.5.8} $$

$$ U_{.1} = b_{.1} \tag{19.5.9} $$

$$ U_{L_2 1} = \text{smallest } U_{.1} \tag{19.5.10} $$

$$ U_{.i} = U_{.i-1} + b_{.i-1} \tag{19.5.11} $$

$$ U_{L_i i} = \text{smallest } U_{.i} \tag{19.5.12} $$

19.5.3 Computational Instructions. We assume that a normalized data matrix is available. Ordinarily, one would not normalize a complete data matrix with a large number of variables if the computations are done with a desk computer. It is assumed, however, that for this particular model a high speed computer is available. It is also assumed that a preliminary computer program is available for transforming the raw data matrix to one whose means are 0 and whose variances are unity.

We begin by considering Eq. (19.5.1). If $\underline{i} = 1$, this is the standardized data matrix and we bypass this equation to get a first approximation to the first factor loading vector.

This is indicated in Eq. (19.5.2). On the basis of some rationale or hypothesis, we may select some particular variable as a satisfactory approximation to one of the factors. If no satisfactory rationale is available, we may arbitrarily begin with the first variable. It is seen, therefore, that the vector given by Eq. (19.5.2) is the correlation of the selected variable $\underline{L}$ with all of the variables, including itself.

Eqs. (19.5.3) through (19.5.8) give the successive cycles required for a particular approximation to the factor loading vector $\underline{b_{.i}}$. We shall discuss this set of computations before indicating generally how we get the first approximation for any particular factor vector following the first.

Eq. (19.5.3) shows the minor product moment of the current approximation to the $\underline{i}$th factor vector. This scalar, $\underline{\gamma}$, indicates the amount of variance accounted for by the $\underline{s}$ approximation to the $\underline{i}$th factor vector.

The $\underline{\beta}$ vector is given in Eq. (19.5.4). This is obtained by constructing first a vector of the cubes of the elements in the current approximation to the $\underline{i}$th factor vector, and subtracting from it the current approximation multiplied by the scalar of Eq. (19.5.3) divided by $\underline{n}$.

Next we compute the product indicated in Eq. (19.5.5), which is the $\underline{i}$th residual data matrix postmultiplied by the vector of Eq. (19.5.4).

Then we calculate the scalar indicated by Eq. (19.5.6) which is the square root of the minor product moment of the vector calculated in Eq. (19.5.5).

We next calculate the current approximation to the factor score vector $\underline{Y}_{.i}$ as indicated in Eq. (19.5.7). This is the vector calculated in Eq. (19.5.5) divided by the scalar calculated in Eq. (19.5.6).

Finally, we calculate the $\underline{s}$+1 approximation to the $\underline{b}_{.i}$ vector by means of Eq. (19.5.8). This is the transpose of the $\underline{i}$th residual data matrix post-multiplied by the factor score vector of Eq. (19.5.7).

We then begin again with Eq. (19.5.3) and repeat the cycle. We continue this set of iterations until either the $\underline{\gamma}$ scalar of Eq. (19.5.3) or the $\underline{\alpha}$ scalar of Eq. (19.5.6) is stabilized to some specified degree.

Then we return to Eq. (19.5.1) to calculate a new residual matrix, which is obtained by subtracting the major product of the stabilized $\underline{Y}$ and $\underline{b}$ vectors of Eqs. (19.5.7) and (19.5.8), respectively, from the current residual matrix.

To get the first approximation to the second varimax factor vector, we consider Eq. (19.5.9). Here we simply equate the $\underline{U}_{.1}$ vector to the first stabilized factor loading vector $\underline{b}_{.1}$.

We then find the algebraically smallest element in the vector of Eq. (19.5.9), as indicated in Eq. (19.5.10).

Next we get the first approximation to the second factor vector by letting $\underline{i}$ = 2 in Eq. (19.5.2).

To get the first approximation to the $\underline{i}$th factor loading vector we consider Eq. (19.5.11). To get $\underline{U}_{.i}$, we add $\underline{U}_{.i-1}$ to the stabilized $\underline{b}_{.i-1}$ varimax vector.

Eq. (19.5.12) indicates the algebraically smallest element in a vector of Eq. (19.5.11). This we designate as in the $\underline{L}_i$ position.

Then we return to Eq. (19.5.2) to get the first approximation to the $\underline{i}$th varimax factor loading vector. This is the transpose of the $\underline{i}$th residual data matrix postmultiplied by the $\underline{L_i}$ column of this residual matrix.

19.5.4 Numerical Example. This numerical example begins with the data matrix given in Table 19.3.2.

Table 19.5.1 shows the first three varimax factor vectors obtained by the successive factor method directly from the data matrix for the case of $\underline{k} = 2$.

Table 19.5.2 gives the varimax factor matrix obtained by the successive factor general method from the data matrix for the case of $\underline{k} = \frac{10}{3}$.

Table 19.5.1 - Successive Factor Varimax Matrix from Data Matrix for $\underline{k} = 2$

0.934	0.091	-0.010
0.934	0.127	0.016
0.891	0.076	0.169
0.093	0.165	0.726
0.044	0.077	0.932
0.096	0.093	0.839
0.236	0.881	0.081
0.240	0.871	0.012
0.300	0.541	0.212

Table 19.5.2 - Successive Factor General Varimax Matrix from Data matrix for $\underline{k} = \frac{10}{3}$

0.943	-0.226	-0.052
0.947	0.320	-0.006
0.884	-0.156	0.136
0.136	-0.050	0.606
0.079	-0.019	0.991
0.136	-0.016	0.700
0.304	0.072	0.124
0.307	0.107	0.069
0.363	0.048	0.225

19.6 Simultaneous Factor Matrix from Data Matrix

19.6.1 Characteristics of the Method. The characteristics of this method have already been fairly well covered in the previous sections. Except for decimal accuracy, it should give essentially the same results as the method discussed in Section 19.4. The calculation of the correlation matrix as such is bypassed, and the multiplications implied by the correlation matrix, that is, the minor product moment of the data matrix, is accomplished at each iteration by two successive multiplications of a matrix by a vector.

The method avoids the accumulation of decimal error resulting from the calculation of residual matrices, such as in Sections 19.3 and 19.5. However, as in Section 19.4, for each approximation one must calculate the basic orthonormal and basic diagonal of a Gramian matrix whose order is equal to the number of factors. Again, this is not a formidable task for high speed computers, since a number of computer programs are already available for computing all of the latent roots and vectors of the Gramian matrix, including the programs in the appendix for Chapter 9.

19.6.2 Computational Equations

19.6.2a Definition of Notations

$\underline{Z}$ is the data matrix with means of 0 and variances of unity.

$\underline{{}_s b}$, $\underline{{}_s b^{(3)}}$, $\underline{{}_s Q}$, and $\underline{{}_s \Delta^2}$ are the same as in Section 19.5.2b.

$\underline{{}_s Y}$ is the $\underline{s}$ approximation to the varimax factor score matrix.

19.6.2b The Equations

$$ {}_1b = Z' Z_{(m)} \tag{19.6.1} $$

$$ {}_s\beta = {}_sb^{(3)} - {}_sb \frac{D_{{}_sb' {}_sb}}{n} \tag{19.6.2} $$

$$ {}_sU = Z \, {}_s\beta \tag{19.6.3} $$

$$ {}_sG = {}_sU' \, {}_sU \tag{19.6.4} $$

$$ {}_sQ \, {}_s\Delta^2 \, {}_sQ' = {}_sG \tag{19.6.5} $$

$$ {}_sG^{-\frac{1}{2}} = {}_sQ \, {}_s\Delta^{-1} \, {}_sQ \tag{19.6.6} $$

$$ {}_sY = {}_sU \, {}_sG^{-\frac{1}{2}} \tag{19.6.7} $$

$$ {}_{s+1}b = Z' \, {}_sY \tag{19.6.8} $$

19.6.3 Computational Instructions. In this model we begin with a standardized data matrix.

We must choose some sort of approximation to the first varimax factor loading matrix. If we have some hypothesis as to a single variable which best measures each of the factors, we can use these variables to begin the computations. In any case, whether we have a rational procedure, or whether we select the first $\underline{m}$ variables where $\underline{m}$ is the number of factors we expect to solve for, we begin with Eq. (19.6.1). The right side of this equation is the transpose of the data matrix postmultiplied by a submatrix made up of $\underline{m}$ vectors out of $\underline{Z}$. These $\underline{m}$ vectors may be rationally or arbitrarily selected. Actually, then, this first approximation to the $\underline{b}$ matrix is simply a matrix of the correlations of the $\underline{m}$ variables with all the variables, including the correlations among themselves.

The general equations for the computations are then indicated by Eqs. (19.6.2) through (19.6.7).

Eq. (19.6.2) gives the computation for the first approximation to the $\underline{\beta}$ matrix, just as in Section 19.4.2. The first term on the right of this equation is a matrix whose elements are the cubes of the corresponding elements of the current approximation to the $\underline{b}$ or varimax factor loading matrix. From this is subtracted the current approximation to the $\underline{b}$ matrix, scaled by a diagonal matrix on the right. This diagonal matrix is made up of the diagonal elements of the minor product moment of the current approximation to the $\underline{b}$ matrix, divided by $\underline{n}$, the number of variables.

The next step is given by Eq. (19.6.3). As indicated on the right hand side of this equation, the data matrix $\underline{Z}$ is postmultiplied by the $\underline{\beta}$ matrix given in Eq. (19.6.2).

The next step is given in Eq. (19.6.4). The matrix $\underline{G}$ is the minor product moment of the matrix calculated in Eq. (19.6.3).

We then calculate the basic structure factors of the matrix $\underline{G}$ given in Eq. (19.6.4), as indicated by the left hand side of Eq. (19.6.5). The computer programs given for Chapter 9 for finding basic structure factors, or eigenvalues and eigenvectors, of symmetric matrices are applicable here.

Next we calculate the $\underline{G}^{-\frac{1}{2}}$ matrix of Eq. (19.6.6). This is the triple product involving the factors obtained from Eq. (19.6.5).

We then calculate the current approximation to the varimax factor score matrix, as indicated in Eq. (19.6.7). This is the matrix of Eq. (19.6.3) postmultiplied by the matrix of Eq. (19.6.6).

Finally, we calculate the next approximation to the $\underline{b}$ or varimax factor loading matrix, as indicated in Eq. (19.6.8). This is simply the transpose of the data matrix $\underline{Z}$ postmultiplied by the factor score matrix given in Eq. (19.6.7).

These computations continue until either the trace of the minor product moment of the current factor loading approximation matrix, or the trace of the $\underline{G}$ matrix in Eq. (19.6.4), reaches some specified degree of stabilization.

19.6.4 Numerical Example. This numerical example begins with the data matrix given in Table 19.3.2.

Table 19.6.1 gives the first three varimax factor vectors obtained by the simultaneous factor matrix method directly from the data matrix for the case of $\underline{k} = 2$.

Table 19.6.2 shows the varimax factor matrix obtained by the simultaneous factor general matrix method directly from the data matrix for the case of $\underline{k} = \frac{10}{3}$.

Table 19.6.1 - Simultaneous Factor Varimax Matrix from Data Matrix for $\underline{k} = 2$

0.932	0.000	0.161
0.932	0.026	0.192
0.893	0.180	0.121
0.082	0.755	0.145
0.040	0.961	0.031
0.089	0.874	0.064
0.181	0.091	0.901
0.185	0.021	0.899
0.269	0.227	0.562

Table 19.6.2 - Simultaneous Factor General Varimax Matrix from Data Matrix for $\underline{k} = \frac{10}{3}$

0.945	0.018	0.136
0.948	0.034	0.147
0.874	0.224	0.066
0.083	1.036	0.087
0.046	0.694	0.012
0.102	0.630	0.093
0.194	0.118	0.972
0.241	0.081	0.641
0.312	0.302	0.446

19.7 Successive Factor General Varimax

As we recall, the general varimax method is similar to the Kaiser varimax method, except that it is based on some even fractional power, greater than unity, of the elements whose variance is maximized, rather than on the squares of these elements.

19.7.1 Computational Equations

19.7.1a Definition of Notation

$\underline{m_1}$ is an integer not less than 1.

$\underline{m_2}$ is an integer not greater than $\underline{m_1}$.

$\underline{{}_s b_{.i}}$ is the $\underline{s}$ approximation to the $\underline{i}$th varimax factor vector.

$\underline{{}_s b_{.i}^{(2k-1)}}$, $\underline{{}_s b_{.i}^{(k-1)}}$, $\underline{{}_s b_{.i}^{(k)}}$ are vectors whose elements are, respectively, the 2$\underline{k}$-1, $\underline{k}$-1, and $\underline{k}$ powers of those in $\underline{{}_s b_{.i}}$.

19.7.1b The Equations

$$k = \frac{2\,m_1}{2\,m_2 - 1} \tag{19.7.1}$$

$${}_s\beta_{.i} = {}_s b_{.i}^{(2k-1)} - {}_s b_{.i}^{(k-1)} \frac{1'\,{}_s b_{.i}^{(k)}}{n} \tag{19.7.2}$$

19.7.2 Computational Instructions. The procedures here are precisely the same as in Sections 19.3 and 19.5, respectively, except that the $\underline{\beta}$ vectors are calculated in a different manner, since the function whose variance we want to maximize is more general than that of the square of the factor loading.

We begin with Eq. (19.7.1). Here we have, instead of the second power, the $\underline{k}$th power of the elements of the factor loading vectors whose variance we wish to maximize. This is expressed as the ratio of twice the sum of a positive integer, divided by twice some other positive integer less 1. The positive integer in the denominator of this equation cannot be greater than that in the numerator.

To define the $\underline{s}$ approximation to the $\underline{\beta}$ vector corresponding to the $\underline{i}$th varimax factor loading vector, we use Eq. (19.7.2). This is the same as Eq. (18.5.1) of Chapter 18. As indicated in that chapter, either we will require tables of logs and exponentials to calculate the powers of the elements of $\underline{b}$ indicated on the right hand side of Eq. (19.7.2), or library functions for the computer program must be available.

19.8 Simultaneous General Varimax

The simultaneous general varimax procedure is similar to procedures described in Sections 19.4 and 19.6, except for the power of the elements in the simple structure matrices whose variances are to be maximized.

19.8.1 Computational Equations

19.8.1a Definition of Notation

$\underline{m_1}$ and $\underline{m_2}$ are the same as in Section 19.7.2a.

$\underline{{}_sb}$ is the $\underline{s}$ approximation to the varimax factor matrix.

$\underline{{}_sb^{(2k-1)}}$, $\underline{{}_sb^{(k)}}$, $\underline{{}_sb^{(k-1)}}$ are matrices whose elements are, respectively, the $2\underline{k}$-1, $\underline{k}$, and $\underline{k}$-1 powers of the corresponding elements in $\underline{{}_sb}$.

$D_{1'_s b^{(k)}}$ is a diagonal matrix whose elements are those of $1'_s b^{(k)}$.

19.8.1b The Equations

$$k = \frac{2\,m_1}{2\,m_2 - 1} \tag{19.8.1}$$

$$_s\beta = {_s b^{(2k-1)}} - {_s b^{(k-1)}} \frac{D_{1'_s b^{(k)}}}{n} \tag{19.8.2}$$

19.8.2 Computational Instructions. The computational instructions are the same as for the method using the correlation matrix, given in Section 19.4, and the method using the data matrix, given in Section 19.6, except for the calculation of the β matrices.

We shall first consider the general case for both the correlation and the data matrices. Again, as in Section 19.7, we begin with Eq. (19.8.1) which gives the value of k as the power of the elements of the simple structure factor loading vector whose variance we wish to maximize. The restrictions on m_1 and m_2 on the right hand side of this equation are the same as the previous ones.

The general equation for the β matrix is now given by Eq. (19.8.2), where the exponents in parentheses for the b matrices indicate that the corresponding elements of the current approximation to the b matrix have been raised to the indicated power. The diagonal matrix on the right of the right hand term of the right side of Eq. (19.8.2) is obtained as follows. We sum the columns of the matrix whose elements are the kth power of the elements in the b matrix, and use the elements of this vector in the diagonal. This diagonal is then divided by n, and the b matrix with elements raised to the

$\underline{k}$ - 1 power is scaled accordingly.

19.9 Mathematical Proofs

19.9.1 Successive Factor Matrix from Correlation Matrix.

From Section 18.3 we have, as the iterative solution for the varimax factor vector $\underline{b_{.i}}$,

$$ {}_s\beta_{.i} = {}_sb^{(3)}_{.i} - {}_sb_{.i}\,\frac{1'\,{}_sb^{(2)}_{.i}}{n} \tag{19.9.1}$$

$$ {}_sU_{.i} = {}_ia'\,{}_s\beta_{.i} \tag{19.9.2}$$

$$ {}_s\alpha_i = \sqrt{{}_sU'_{.i}\,{}_sU_{.i}} \tag{19.9.3}$$

$$ {}_sH_{.i} = \frac{{}_sU_{.i}}{{}_s\alpha_i} \tag{19.9.4}$$

$$ {}_{s+1}b_{.i} = {}_ia\,{}_sH_{.i} \tag{19.9.5}$$

where

$$ {}_ia = {}_{i-1}a - b_{.(i-1)}\,H'_{.(i-1)} \tag{19.9.6}$$

From Eq. (19.9.2)

$$ {}_ia\,{}_ia'\,{}_s\beta_{.i} = {}_ia\,{}_sU_{.1} \tag{19.9.7}$$

From Eqs. (19.9.5) and (19.9.6)

$$ {}_ia\,{}_ia' = {}_{(i-1)}a\,{}_{(i-1)}a' - b_{.(i-1)}\,b'_{.(i-1)} \tag{19.9.8}$$

From Eqs. (19.9.4), (19.9.5), and (19.9.7)

$$ {}_ia\,{}_ia'\,{}_s\beta_{.i} = {}_sb_{.i}\,{}_s\alpha_i \tag{19.9.9}$$

From Eq. (19.9.3)

$$ {}_s\alpha_i = \sqrt{{}_s\beta'_{.i}\ {}_i a\ {}_i a'\ {}_s\beta_{.i}} \tag{19.9.10}$$

We now let

$$ {}_1 R = {}_1 a\ {}_1 a' \tag{19.9.11}$$

where ${}_{\underline{1}} a$ is the factor loading matrix.

From Eqs. (19.9.8) and (19.9.11)

$$ {}_2 R = {}_1 R - b_{.1}\ b'_{.1} \tag{19.9.12}$$

or, in general,

$$ {}_{i+1} R = {}_i R - b_{.i}\ b'_{.i} \tag{19.9.13}$$

From Eqs. (19.9.8) and (19.9.13)

$$ {}_{i+1} a\ {}_{i+1} a' = {}_{i+1} R \tag{19.9.14}$$

From Eqs. (19.9.9), (19.9.10), and (19.9.14)

$$ {}_{s+1} b_{.i} = \frac{{}_i R\ {}_s\beta_{.i}}{\sqrt{{}_s\beta'_{.i}\ {}_i R\ {}_s\beta_{.i}}} \tag{19.9.15}$$

From Eqs. (19.9.1), (19.9.13), and (19.9.15) we can solve iteratively for $\underline{b_{.i}}$. From Eq. (19.9.15) we see that, for any iteration $\underline{s}$, $\underline{{}_s b_{.i}}$ is a rank reduction solution for Eq. (19.9.13), irrespective of how well the solution has stabilized.

We let

$$ {}_s\gamma_i = {}_s b'_{.i}\ {}_s b_{.i} \tag{19.9.16}$$

Then from Eqs. (19.9.1) and (19.9.16)

$$_s\beta_{.i} = {}_s b^{(3)}_{.i} - {}_s b_{.i} \frac{_s\gamma_i}{n} \tag{19.9.17}$$

Let

$$_s W_{.i} = {}_i R \; {}_s\beta_{.i} \tag{19.9.18}$$

From Eqs. (19.9.10), (19.9.14), and (19.9.18)

$$_s\alpha_i = \sqrt{_s W'_{.i} \; {}_s\beta_{.i}} \tag{19.9.19}$$

From Eqs. (19.9.15), (19.9.18), and (19.9.19)

$$_{s+1} b_{.i} = \frac{_s W_{.i}}{_s\alpha_i} \tag{19.9.20}$$

The computational equations then for the $b_{.i}$ are given by Eqs. (19.9.16) through (19.9.20), and the successive $_i R$'s are calculated from Eq. (19.9.13). The variance reduction in the $_i R$ matrix accounted for by the s approximation to $_s b_{.i}$ is obviously given by $_s\gamma_i$ in Eq. (19.9.16).

19.9.2 Simultaneous Factor Matrix from Correlation Matrix

From Section 18.4.2 we have, as the s approximation to the simultaneous varimax factor matrix,

$$_s\beta = {}_s b^{(3)} - {}_s b \frac{D_{\,{}_s b' \, {}_s b}}{n} \tag{19.9.21}$$

$$a' \, {}_s\beta = {}_s C \tag{19.9.22}$$

$$_s C' \, {}_s C = {}_s Q \, {}_s\Delta^2 \, {}_s Q' \tag{19.9.23}$$

$$_s H = {}_s C \, {}_s Q \, {}_s\Delta^{-1} \, {}_s Q' \tag{19.9.24}$$

$$_{s+1} b = a \, {}_s H \tag{19.9.25}$$

From Eqs. (19.9.22) and (19.9.23)

$$ {}_s\beta' \; a \; a' \; {}_s\beta \;=\; {}_sQ \; {}_s\Delta^2 \; {}_sQ' \tag{19.9.26} $$

From Eqs. (19.9.22), (19.9.24), and (19.9.25)

$$ {}_{s+1}b \;=\; a \; a' \; {}_s\beta \; {}_sQ \; {}_s\Delta^{-1} \; {}_sQ' \tag{19.9.27} $$

We let

$$ a \; a' \;=\; R \tag{19.9.28} $$

From Eqs. (19.9.26) and (19.9.28)

$$ {}_sQ \; {}_s\Delta^2 \; {}_sQ' \;=\; {}_s\beta' \; R \; {}_s\beta \tag{19.9.29} $$

From Eqs. (19.9.27) and (19.9.28)

$$ {}_{s+1}b \;=\; R \; {}_s\beta \; {}_sQ \; {}_s\Delta^{-1} \; {}_sQ' \tag{19.9.30} $$

Let

$$ {}_sM \;=\; R \; {}_s\beta \tag{19.9.31} $$

$$ {}_sG \;=\; {}_s\beta' \; {}_sM \tag{19.9.32} $$

From Eqs. (19.9.29) and (19.9.32)

$$ {}_sQ \; {}_s\Delta^2 \; {}_sQ' \;=\; {}_sG \tag{19.9.33} $$

From Eqs. (19.9.30) and (19.9.31)

$$ {}_{s+1}b \;=\; {}_sM \; {}_sQ \; {}_s\Delta^{-1} \; {}_sQ' \tag{19.9.34} $$

Then Eqs. (19.9.21) and (19.9.31) through (19.9.34) constitute the computational equations for calculating the $\underline{s}$ approximation to the $\underline{b}$ matrix. That any approximation ${}_{s}b$ of width m is a solution for

$$_{m}R = R - {}_{s}b \; {}_{s}b' \tag{19.9.35}$$

such that the rank of ${}_{m}R$ is m less than the rank of R, can be readily shown as follows. Dropping the prescripts, we have from Eqs. (19.9.29) and (19.9.30),

$$b = R\,\beta\,(\beta'\,R\,\beta)^{-\frac{1}{2}} \tag{19.9.36}$$

Substituting Eq. (19.9.36) in Eq. (19.9.35)

$$_{m}R = R - R\,\beta\,(\beta'\,R\,\beta)^{-1}\,\beta'\,R \tag{19.9.37}$$

which is, of course, the rank reduction form.

The iterations may proceed until $\mathrm{tr}\;{}_{s}G$ converges to the desired degree of decimal accuracy.

19.9.3 Successive Factor Varimax from Data Matrix

Given the data matrix Z such that

$$R = Z'\,Z \tag{19.9.38}$$

The successive factor varimax from the correlation matrix, according to Section 19.3.2, is given by the set of equations

$$_{s}\gamma_{i} = {}_{s}b'_{.i}\;{}_{s}b_{.i} \tag{19.9.39}$$

$$_{s}\beta_{.i} = {}_{s}b^{(3)}_{.i} - {}_{s}b_{.i}\,\frac{{}_{s}\gamma_{i}}{n} \tag{19.9.40}$$

$$_{s}W_{.i} = {}_{i}R\;{}_{s}\beta_{.i} \tag{19.9.41}$$

$$ {}_s\alpha_i = \sqrt{{}_sW'_{.i}\ {}_s\beta_{.i}} \qquad (19.9.42) $$

$$ {}_{s+1}b_{.i} = \frac{{}_sW_{.i}}{{}_s\alpha_i} \qquad (19.9.43) $$

and

$$ {}_{i+1}R = {}_iR - b_{.i}\ b'_{.i} \qquad (19.9.44) $$

From Eqs. (19.9.38), (19.9.41), (19.9.42) and (19.9.43)

$$ {}_{s+1}b_{.i} = \frac{Z'\ Z\ {}_s\beta_{.i}}{\sqrt{{}_s\beta'_{.i}\ Z'\ Z\ {}_s\beta_{.i}}} \qquad (19.9.45) $$

Let

$$ {}_sU_{.i} = Z\ {}_s\beta_{.i} \qquad (19.9.46) $$

$$ {}_sY_{.i} = \frac{{}_sU_{.i}}{\sqrt{{}_sU'_{.i}\ {}_sU_{.i}}} \qquad (19.9.47) $$

From Eqs. (19.9.45), (19.9.46), and (19.9.47)

$$ {}_{s+1}b_{.i} = Z'\ {}_sY_{.i} \qquad (19.9.48) $$

Consider now the residual matrix

$$ {}_2Z = {}_1Z - Y_{.1}\ b'_{.1} \qquad (19.9.49) $$

From Eqs. (19.9.48) and (19.9.49)

$$ {}_2Z = {}_1Z - Y_{.1}\ Y'_{.1}\ {}_1Z \qquad (19.9.50) $$

or

$$ {}_2Z = (I - Y_{.1}\ Y'_{.1})\ {}_1Z \qquad (19.9.51) $$

From Eqs. (19.9.47) and (19.9.51)

$$ {}_2Z' \; {}_2Z = {}_1Z' \, (I - Y_{.1} \, Y'_{.1}) \, {}_1Z \tag{19.9.52} $$

From Eqs. (19.9.38), (19.9.44), (19.9.48), and (19.9.52)

$$ {}_2Z' \; {}_2Z = {}_1R - b_{.1} \, b'_{.1} \tag{19.9.53} $$

or from Eqs. (19.9.38) and (19.9.53)

$$ {}_2Z' \; {}_2Z = {}_2R \tag{19.9.54} $$

In general, if

$$ {}_{i+1}Z = {}_iZ - Y_{.i} \, b'_{.i} \tag{19.9.55} $$

then

$$ {}_{i+1}Z' \; {}_{i+1}Z = {}_{i+1}R \tag{19.9.56} $$

Eqs. (19.9.39), (19.9.40), (19.9.46), (19.9.47), (19.9.48), and (19.9.55) may therefore be used to calculate the successive varimax factor vectors directly from the standard score matrix. Since Eq. (19.9.44) is a rank reduction form, Eq. (19.9.56) shows that Eq. (19.9.55) is also a rank reduction form for any approximation $\underline{s}$ to $\underline{Y_{.i}}$ and $\underline{b_{.i}}$.

Consider then

$$ Y = (Y_{.1} \cdots Y_{.m}) \tag{19.9.57} $$

$$ b = (b_{.1} \cdots b_{.m}) \tag{19.9.58} $$

Then in

$$_{m}Z = Z - Y\, b' \tag{19.9.59}$$

the rank of $_{m}Z$ is m less than the rank of Z. Also, it should now be obvious that Y is the factor score matrix corresponding to the factor loading matrix b.

19.9.4 Simultaneous Varimax from the Data Matrix

Consider again the data matrix Z, such that

$$R = Z'\, Z \tag{19.9.60}$$

The simultaneous varimax matrix from the correlation matrix, according to Section 19.4, is given by the set of equations

$$_{s}\beta = {}_{s}b^{(3)} - {}_{s}b \, \frac{D_{{}_{s}b'\,{}_{s}b}}{n} \tag{19.9.61}$$

$$_{s}M = R\, {}_{s}\beta \tag{19.9.62}$$

$$_{s}G = {}_{s}\beta'\, {}_{s}M \tag{19.9.63}$$

$$_{s+1}b = {}_{s}M\, {}_{s}G^{-\frac{1}{2}} \tag{19.9.64}$$

From Eqs. (19.9.60) and (19.9.62)

$$_{s}M = Z'\, Z\, {}_{s}\beta \tag{19.9.65}$$

From Eqs. (19.9.60), (19.9.62), and (19.9.63)

$$_{s}G = {}_{s}\beta'\, Z'\, Z\, {}_{s}\beta \tag{19.9.66}$$

From Eqs. (19.9.62) through (19.9.66)

$$ {}_{s+1}b = Z' \, Z \, {}_s\beta \, ({}_s\beta \, Z' \, Z \, {}_s\beta)^{-\frac{1}{2}} \tag{19.9.67} $$

Let

$$ Z \, {}_s\beta = {}_sU \tag{19.9.68} $$

Let

$$ {}_sU = {}_sP \, {}_s\Delta' \, {}_sQ' \tag{19.9.69} $$

From Eqs. (19.9.66) and (19.9.68)

$$ {}_sG = {}_sU' \, {}_sU \tag{19.9.70} $$

From (19.9.60) and (19.9.70)

$$ {}_sG = {}_sQ \, {}_s\Delta^2 \, {}_sQ' \tag{19.9.71} $$

Let

$$ {}_sY = {}_sU \, {}_sG^{-\frac{1}{2}} \tag{19.9.72} $$

From Eqs. (19.9.71) and (19.9.72)

$$ {}_sY = {}_sU \, {}_sQ \, {}_s\Delta^{-1} \, {}_sQ' \tag{19.9.73} $$

From Eqs. (19.6.67) through (19.9.73)

$$ {}_{s+1}b = Z' \, {}_sY \tag{19.9.74} $$

We can then solve for successive approximations to $\underline{b}$ and $\underline{Y}$ by Eqs. (19.9.61), (19.9.68), (19.9.70), (19.9.71), and (19.9.73).

From Eqs. (19.9.69) and (19.9.73)

$$_sY = {_sP}\,{_sQ'} \tag{19.9.75}$$

hence

$$_sY'\,{_sY} = I \tag{19.9.76}$$

Let

$$_mZ = Z - Y\,b' \tag{19.9.77}$$

From Eqs. (19.9.74) and (19.9.77)

$$_mZ = Z - Y\,Y'\,Z \tag{19.9.78}$$

or

$$_mZ = (I - Y\,Y')\,Z \tag{19.9.79}$$

From Eqs. (19.9.60), (19.9.74), (19.9.76), and (19.9.79)

$$_mZ'\,{_mZ} = R - {_sb}\,{_sb'} \tag{19.9.80}$$

hence

$$_mR = R - {_sb}\,{_sb'} \tag{19.9.81}$$

In Section 19.9.2 we proved the rank of $_mR$ is m less than the rank of R. Hence the rank of $_mZ$ in Eq. (19.9.77) is m less than the rank of Z for any approximation s.

19.9.5 Successive General Varimax Vectors

The direct solutions for the successive factor general varimax differ from the solutions which maximize the variance of squared factor loadings only in the calculation of the ${}_s\beta_{.i}$ vectors.

Let

$$k = \frac{2\ m_1}{2\ m_2 - 1} \tag{19.9.82}$$

where

$$m_1 \geq m_2 \tag{19.9.83}$$

and m_1 and m_2 are both integers greater than 0.

Then

$${}_s\beta_{.i} = {}_s b_{.i}^{(2k-1)} - {}_s b_{.i}^{(k-1)} \frac{1' \, {}_s b_{.i}^{(k)}}{n} \tag{19.9.84}$$

From Eq. (19.9.82)

$$2\ k - 1 = \frac{2\ (2\ m_1 - m_2) + 1}{2\ m_2 - 1} \tag{19.9.85}$$

$$k - 1 = \frac{2\ (m_1 - m_2) + 1}{2\ m_2 - 1} \tag{19.9.86}$$

If we wish to calculate the successive general varimax vectors directly from the correlation matrix, we use the same equations as in Section 19.3.2, with the exception of the ${}_s\beta_{.i}$ vector indicated in Eq. (19.9.84).

If we wish to calculate the successive general varimax vectors directly from the data matrix, we use the same equations as in 19.5.2, with the exception of the ${}_s\beta_{.i}$ vector which is now given by Eq. (19.9.84).

19.9.6 Simultaneous General Varimax Factor Matrix

The rationale for the simultaneous general varimax directly from the correlation or data matrices is the same as for the special case of the squared factor loadings, except that the ${}_s\beta$ matrix is different for values of k other than 2. The general expression for ${}_s\beta$ is given by

$$ {}_s\beta = {}_sb^{(2k-1)} - {}_sb^{(k-1)} D_{1'_s b^{(k)}} \tag{19.9.87} $$

where $D_{1'_s b}$ is a diagonal of the vector of column sums of the ${}_sb^{(k)}$ matrix.

CHAPTER 20

FACTOR SCORE MATRICES

20.1 Introduction

Over the past 40 years a vast amount of attention has been given to the factor analysis of correlation matrices. In this book we have already devoted a large number of chapters to various methods for getting factor loading matrices from correlation matrices. We saw in Chapter 4 how we may view the general factor analysis problem as one of approximating a data matrix by another of lower rank. We saw that the problem viewed in this way is one of finding two basic matrices whose major product is in some sense a satisfactory approximation to the data matrix. The factors of this major product have a common order much less than the smaller order of the data matrix, and therefore the rank of the product is equal to the common order of its factors.

Again in Chapter 4 we saw how we may regard the postfactor of this product as the transpose of the factor loading matrix. We also saw how we may regard the prefactor as the factor score matrix. Therefore, the factor score matrix postmultiplied by the transpose of the factor loading matrix yields the lower rank approximation to the data matrix. In this formulation of the problem, the communality problem does not appear. There seems to be no clear justification for considering the communality concept when we view factor analysis, not as a method of factoring the correlation or covariance matrix, but rather as one of factoring the data matrix. Guttman (1955) has discussed an interesting exception, which is, however, beyond the scope of this book.

In any case, with all of the attention given to the solutions for factor loading matrices, very little has been devoted to the derivation of factor score matrices. This is especially curious inasmuch as the scientific, logical, and philosophical status of the factor score matrix would appear to be at least as respectable as that of the factor loading matrix. Some would argue, of course, that the factor loading matrix is of more fundamental importance because it enables us to identify or define the fundamental variables of a particular discipline. Certainly there is something to be said for this point of view, if one regards the major objective of the simple structure or transformation technique as one of finding factor loadings which are relatively invariant from one sample of entities and attributes to another.

From a philosophical point of view, these considerations may justify greater interest in the factor loading matrix. However, from a purely formal point of view, considering only the mathematics involved, there is no reason to be more interested in the factor loading matrix than in the factor score matrix. Considering the model in which the major product of these two matrices approximates the data matrix, there is actually nothing in the mathematics of the model which would in some sense give higher status to the postfactor than to the prefactor.

This statement is even more cogent if we recall Chapters 13 and 15 dealing with linear transformations involving both scale and origin, which may be applied to both the right and the left hand sides of the data matrix. We have seen that these operations can materially affect the results of a factor analysis, and that the traditional practice of metricizing the data matrix by attributes rather than by entities, is more or less arbitrary. In any case, the problem of finding the left hand factor of the matrix product which purports to approximate the data matrix appears to merit considerably more attention than it has received in the past.

But even the attention which the problem has received seems to have caused as much confusion as clarification for the central issues involved. These are actually rather simple, if one does not become unduly involved with the red herrings of the communality and specificity problems. One need only examine the bewildering, even if at times ingenious, traditional discussions of the factor measurement problem to realize that they have often strayed far from the solid ground of the data matrix.

Some investigators have argued that factor scores give no more information than do the measures from which they are derived, and that therefore, at best these scores are of more theoretical than practical interest. Unfortunately, these investigators have asked the wrong question. Instead of asking whether the factor scores give more information than the original data scores, they should have asked whether the original data scores give more information than the factor scores. If one asks this latter question, he may conclude that the data matrix may yield not only relevant or systematic information, but also random or unreliable information.

One may then regard the lower rank data matrix approximation model as a procedure for eliminating from the data matrix random or irrelevant variance. Horst (1941) has utilized the factor score matrix to reduce the effect of this variance. The method has never received wide attention. Leiman (1951) has applied this conception of factor analysis objectives to experimental data. He has found that, by the use of lower rank approximation matrices to data matrices, one may obtain multiple regression parameters which hold up better on cross-validation than when the data matrices are employed directly in the traditional methods. A much more extensive application of factor score matrices and the lower rank approximation model has been made by Burket (1964). In his work it is clear that for prediction purposes the lower rank approximation procedures have a definite advantage over the conventional multiple regression procedures. This application of the factor score concept will be considered in more detail in Chapter 23. Here we will present various types of methods for calculating factor score matrices.

20.2 Kinds of Factor Score Solutions

We may classify the various kinds of solutions for factor score matrices to correspond with the procedures for getting factor loading matrices which we have discussed in the previous chapters. The solutions will be presented under the headings of the centroid factor score matrix, the multiple group factor score matrix, the principal axis factor score matrix, the least square factor score matrix, and the image analysis factor score matrix.

20.2.1 The Centroid Factor Score Matrix. The calculation of a centroid factor score matrix directly from the data matrix has already been explicitly considered in Chapter 12. In this process, we arrive successively at factor loading vectors and factor score vectors. However, in this chapter we shall present the calculation of a factor score matrix based on a previous calculation of the centroid factor loading matrix. This factor loading matrix, together with a matrix of sign vectors, yields a transformation matrix which, when applied to the data matrix, gives the centroid factor score matrix.

20.2.2 The Multiple Group Factor Score Matrix. We shall see that by the use of a binary grouping matrix we can also calculate a multiple group factor score matrix directly from the data matrix. As a matter of fact, it is easier to use the multiple group method directly on the data matrix than it is to use the centroid method, because in the centroid method we must have a matrix of sign vectors, and this ordinarily becomes available only with a successive factor solution. In the centroid method, it will be recalled that residual matrices are calculated, and with each residual matrix one iterates to the optimal sign vector.

On the other hand, in the case of the multiple group method, one presumably has an a priori binary grouping matrix. For this reason one need not go through the actual calculation of the multiple group factor loading matrix before calculating the multiple group factor score matrix.

20.2.3 The Principal Axis Factor Score Matrix. Perhaps the simplest and most elegant of all of the methods for getting factor score matrices is the basic structure method. As we have seen in Chapter 4, the principal axis or basic structure type solution yields the least square approximation to the data matrix for any specified rank of the approximation matrix. We have seen that the principal axis or basic structure factor loading matrix is both a rank reduction solution and a least square approximation to the correlation or covariance matrix. It is easy to show that the first m vectors of the left basic orthonormal matrix of the data matrix yield precisely the principal axis factor score matrix, and that this is an orthogonal matrix.

20.2.4 The Least Square Factor Score Matrix. Having given some arbitrary factor loading matrix, whether centroid, multiple group, or principal axis, we may wish to determine that factor score matrix which, when postmultiplied by the transpose of the factor loading matrix, yields a product which is the least square approximation to the data matrix. This means that the sums of squares of elements of the residual matrix will be a minimum. This will in general be true, as we have seen, for the principal axis method. We can also find factor score matrices for the centroid and the multiple group methods which are least square solutions to the data matrix. As a matter of fact, for any arbitrary basic matrix of width equal to the number of attributes and of height equal to the number of factors, we can find what particular vertical matrix, postmultiplied by the transpose of the arbitrary factor matrix, yields a product which is the best approximation to the data matrix in the least square sense.

20.2.5 The Image Analysis Factor Score Matrix. To our knowledge, no detailed analysis for the calculation of the factor score matrix from the image type factor loading matrix has been previously presented. Harris (1962) has given an interesting theoretical analysis of this problem in a recent paper. The image analysis approach implies a transformation of the data matrix. It is therefore of interest to see what procedures are appropriate in the calculation of factor score matrices based on these image factor loading matrices and the transformed data matrix.

20.3 The Centroid Factor Score Matrix

20.3.1 Characteristics of the Method. We have indicated in the previous section that the centroid factor loading matrix may be used in the solution of a factor score matrix, such that the major product of the two will give the best least square approximation to the data matrix. We shall, however, consider here only a particular type of centroid factor score solution. In this solution the factor score matrix is an orthogonal matrix.

It has been repeatedly said or implied that the centroid solution yields orthogonal factors. But like much of the discussion about orthogonal factors, the definition of orthogonal factors has been vague. We shall insist on using the term orthogonal only with respect to vectors. We shall insist that the term orthogonality is not useful unless considered in this connection. By saying that two vectors are orthogonal, we simply mean that their minor product is 0.

The solution for the centroid factor score matrix which we shall consider does yield factor scores such that the minor product of any pair of factor score vectors taken from the factor score matrix will be 0.

A further characteristic of this method is that the solution is a rank reduction solution. That is, the solution is such that, when the major product of the factor score matrix by the factor loading matrix is subtracted from the data matrix, the residual matrix is of rank equal to the rank of the data matrix less the number of factors.

This solution, like all of the solutions for factor score matrices which we shall consider, involves no iterative procedures. In this respect it is relatively simple and straightforward computationally.

The methods outlined in all of the procedures in this chapter are concerned particularly with the calculation of transformation matrices which may be applied to data matrices to convert them to factor score matrices.

Therefore, in actual practice it may be desirable to use these transformation matrices on data matrices other than those from which the factor loading matrices were calculated. When this is the case, the resulting factor score matrices cannot be expected to exhibit precisely the same characteristics as when these transformation matrices are applied to the original data matrix.

20.3.2 Computational Equations

20.3.2a Definition of Notation

$\underline{a}$ is the centroid factor loading matrix.

$\underline{L}$ is the matrix of sign vectors.

$\underline{Y}$ is the centroid factor score matrix.

$\underline{Z}$ is the normalized data matrix.

20.3.2b The Equations

$$t' = a' L \tag{20.3.1}$$

$$B = L\, t'^{-1} \tag{20.3.2}$$

$$Y = Z\, B \tag{20.3.3}$$

20.3.3 Computational Instructions. We begin with the centroid factor loading matrix $\underline{a}$. We also have given the matrix of sign vectors used in the calculation of the centroid factor loading matrix, which we designate as $\underline{L}$.

We calculate the upper triangular matrix as indicated in Eq. (20.3.1).

This is the transpose of the factor loading matrix postmultiplied by the sign matrix.

Next we calculate Eq. (20.3.2), which is the matrix of sign vectors postmultiplied by the inverse of the triangular matrix in Eq. (20.3.1).

We now calculate the centroid factor score matrix, as indicated in Eq. (20.3.3). This is the data matrix postmultiplied by the transformation matrix of Eq. (20.3.2).

20.3.4 Numerical Example. We begin with the data matrix used in Chapter 19, whose corresponding correlation matrix is the same used in previous chapters. The data matrix is repeated for convenient reference in Table 20.3.1.

Table 20.3.2 gives the centroid factor loading matrix for three factors, calculated from the correlation matrix.

Table 20.3.3 gives the matrix of sign vectors by rows, used in calculating the centroid matrix, and subsequently in the calculation of the factor score matrix.

Table 20.3.4 gives the centroid factor score matrix for three factors, calculated by means of Eqs. (20.3.1) through (20.3.4).

Table 20.3.5 gives the minor product moment of the factor score matrix. As proved in Section 20.8.1, this should be an identity matrix, which it is within limits of rounding error.

Table 20.3.1 - Normalized Deviation Data Matrix

0.128	0.181	0.421	0.506	0.857	0.746	0.280	0.178	0.246
0.764	0.740	0.563	-0.387	-0.293	-0.202	0.261	0.281	0.043
-0.030	-0.046	0.014	0.147	-0.109	-0.135	0.640	0.682	0.661
-0.280	-0.351	-0.326	-0.023	-0.109	-0.186	0.083	0.091	-0.654
-0.336	-0.306	-0.429	-0.542	-0.006	-0.153	-0.428	0.056	-0.124
-0.276	-0.324	-0.271	-0.370	-0.225	0.035	0.129	-0.446	-0.124
0.057	-0.070	-0.016	0.006	0.152	-0.441	-0.166	-0.354	-0.033
-0.010	-0.140	0.326	-0.004	-0.258	-0.091	-0.410	-0.200	-0.101
-0.303	0.227	-0.014	0.029	-0.102	-0.125	-0.086	-0.106	-0.074
0.086	0.057	-0.003	0.161	0.073	0.134	-0.082	-0.067	-0.009
0.164	0.106	-0.124	0.234	-0.002	0.227	-0.072	0.025	0.156
0.036	-0.074	-0.142	0.242	0.023	0.192	-0.148	-0.119	0.011

Table 20.3.2 - Centroid Factor Loading Matrix by Rows for Three Factors

0.659	0.684	0.730	0.617	0.542	0.588	0.637	0.606	0.708
0.457	0.452	0.283	-0.601	-0.709	-0.663	0.271	0.346	0.164
0.456	0.439	0.451	-0.209	0.094	0.116	-0.479	-0.488	-0.378

Table 20.3.3 - Sign Matrix by Rows for Three Factors

1.	1.	1.	1.	1.	1.	1.	1.	1.
1.	1.	1.	-1.	-1.	-1.	1.	1.	1.
1.	1.	1.	-1.	1.	1.	-1.	-1.	-1.

Table 20.3.4 - Normalized Centroid Factor Score Matrix by Columns

0.614	-0.525	0.175
0.307	0.719	0.462
0.316	0.329	-0.810
-0.304	-0.108	-0.191
-0.393	0.007	0.019
-0.324	-0.003	-0.015
-0.150	0.011	0.105
-0.154	0.043	0.211
-0.096	0.015	-0.004
0.061	-0.133	0.083
0.124	-0.123	-0.031
0.004	-0.228	-0.011

Table 20.3.5 - Minor Product Moment of Centroid Factor Score Matrix

0.997	-0.001	0.002
-0.001	0.999	0.000
0.002	0.000	1.001

20.4 The Multiple Group Factor Score Matrix

20.4.1 Characteristics of the Solution. The solution for the multiple group factor scores, as already indicated, does not first require the calculation of the multiple group factor loading matrix. If we have the binary grouping matrix, it can be applied directly to the correlation matrix to yield a transformation matrix which, when applied to the data matrix, gives the factor score matrix. As in the case of the centroid method, this solution is not a least square solution in the sense that the major product of the factor score and factor loading matrices gives the best least square fit to the data matrix. However, it does yield a factor score matrix which is orthonormal and of rank reduction form.

20.4.2 Computational Equations

20.4.2a Definition of Notation

$\underline{R}$ is the correlation matrix.

$\underline{f}$ is a binary grouping matrix.

$\underline{t}$ is a lower triangular matrix.

$\underline{Y}$ is the multiple group factor score matrix.

$\underline{Z}$ is the normalized data matrix.

20.4.2b The Equations

$$G = f' R f \qquad (20.4.1)$$

$$t\, t' = G \qquad (20.4.2)$$

$$B = f\, t'^{-1} \tag{20.4.3}$$

$$Y = Z\, B \tag{20.4.4}$$

20.4.3 Computational Instructions. We assume that the correlation matrix $\underline{R}$ and a binary grouping matrix $\underline{f}$ are given. We then calculate the matrix in Eq. (20.4.1). This is the correlation matrix postmultiplied by the binary grouping matrix and premultiplied by its transpose.

Next we calculate the triangular factors of the matrix of Eq. (20.4.1), as indicated in Eq. (20.4.2).

In Eq. (20.4.3) we postmultiply the binary grouping matrix by the inverse of the upper triangular factor in Eq. (20.4.2).

The multiple group factor score matrix is indicated in Eq. (20.4.4). This is the data matrix postmultiplied by the transformation matrix of Eq. (20.4.3). The minor product moment of this matrix is shown in Section 20.8.2 to be the identity matrix.

20.4.4 Numerical Example. We use the same data matrix and correlation matrix as in the previous section.

Table 20.4.1 gives the binary grouping matrix by rows for three factors.

Table 20.4.2 gives the normalized multiple group factor score matrix by rows for three factors. The matrix was calculated by means of Eqs. (20.4.1) through (20.4.4).

Table 20.4.3 gives the minor product moment of the multiple group factor score matrix. Within rounding error this is an identity matrix, as it should be according to Section 20.8.2. In this sense the multiple group factors may be said to be orthogonal, but only if the factor scores are calculated in this manner.

The same may also be said for the centroid factors when the factor score matrix is calculated according to Section 20.3.

Table 20.4.1 - Binary Grouping Matrix by Rows for Three Factors

1.	1.	1.	0.	0.	0.	0.	0.	0.
0.	0.	0.	1.	1.	1.	0.	0.	0.
0.	0.	0.	0.	0.	0.	1.	1.	1.

Table 20.4.2 - Normalized Multiple Group Factor Score Matrix for Three Factors

0.262	0.767	0.007
0.743	-0.470	0.022
-0.022	-0.033	0.902
-0.344	-0.062	-0.037
-0.385	-0.203	0.009
-0.313	-0.161	-0.010
-0.010	-0.107	-0.216
0.063	-0.147	-0.312
-0.032	-0.071	-0.086
0.050	0.133	-0.126
0.052	0.168	-0.016
-0.065	0.187	-0.128

Table 20.4.3 - Minor Product Moment of Multiple Group Factor Score Matrix

1.000	0.000	-0.003
0.000	1.000	-0.001
-0.003	-0.001	1.000

20.5 The Principal Axis Factor Score Matrix

20.5.1 Characteristics of the Method. As indicated in the previous discussion on kinds of methods, the principal axis solution is the simplest of the methods, if the basic structure or principal axis factor loading matrix is already available. One of the most important advantages of the principal axis method is that it gives, at the same time, a least square, a rank reduction, and an orthogonal solution for the factor score matrix.

20.5.2 Computational Equations

20.5.2a Definition of Notation

$\underline{a}$ is the principal axis factor loading matrix.

$\underline{\delta}$ is the basic diagonal of the correlation matrix.

$\underline{Y}$ is the principal axis factor score matrix.

$\underline{Z}$ is the normalized data matrix.

20.5.2b The Equations

$$B = a\,\delta^{-1} \qquad (20.5.1)$$

$$Y = Z\,B \qquad (20.5.2)$$

20.5.3 Computational Instructions. The computational instructions for the principal axis factor score matrix are very simple. We begin with the factor loading matrix $\underline{a}$ and the basic diagonal $\underline{\delta}$. Eq. (20.5.1), then, directly gives the transformation matrix. This is simply the factor loading matrix postmultiplied by the inverse of the $\underline{\delta}$ diagonal.

The factor score matrix is then calculated in the usual manner, as indicated in Eq. (20.5.2).

Hotelling (1933) published this solution for the factor score matrix. However, it does not seem to be well known and has not been used extensively.

21.5.4 Numerical Example. In this example we use the same data and correlation matrices as in the previous sections.

Table 20.5.1 gives the first three basic diagonals of the correlation matrix, as found in early chapters giving basic structure solutions.

Table 20.5.2 gives the first three principle axis factor vectors by rows, as found in previous chapters, for example, Chapters 7 and 8.

Table 20.5.3 gives the normalized principal axis factor score matrix for three factors, as calculated from Eqs. (20.5.1) and (20.5.2).

Table 20.5.4 gives the minor product moment of the principal axis factor score matrix. This is the identity matrix to within rounding error.

Table 20.5.1 - First Three Basic Diagonals of Correlation Matrix

3.749	2.050	1.331

Table 20.5.2 - Principal Axis Factor Loading Matrix by Rows for Three Factors

0.717	0.740	0.773	0.556	0.463	0.518	0.640	0.615	0.715
0.493	0.478	0.296	-0.649	-0.744	-0.694	0.080	0.166	-0.034
0.350	0.322	0.406	0.068	0.181	0.188	-0.588	-0.621	-0.369

Table 20.5.3 - Normalized Principal Axis Factor Score Matrix for Three Factors

0.555	-0.569	0.179
0.386	0.767	0.205
0.325	0.092	-0.825
-0.328	-0.065	-0.198
-0.403	0.001	-0.146
-0.325	-0.023	-0.093
-0.140	0.053	0.200
-0.129	0.107	0.317
-0.088	0.037	0.049
0.052	-0.098	0.143
0.114	-0.107	0.052
-0.014	-0.195	0.109

Table 20.5.4 - Minor Product Moment of Principal Axis Factor Score Matrix

0.998	-0.001	0.002
-0.001	1.000	0.000
0.002	0.000	1.002

20.6 The Least Square Factor Score Matrix

20.6.1 Characteristics of the Method. This method is not mutually exclusive of those previously considered. It may be applied to any factor loading matrix such as the centroid, the multiple group, or the principal axis. When applied to the principal axis factor matrix it yields precisely the solution given in the preceding section. The least square solution yields a factor score matrix such that, when the major product of this matrix and the factor loading matrix is subtracted from the data matrix, the sum of squares of elements in the residual matrix is a minimum. This solution, as all least square solutions in general, can be shown to be a rank reduction solution. The left arbitrary multiplier, however, is somewhat more involved than in other methods, as can be seen from Section 20.9.4. In general, also, the computations are somewhat more involved than they are for the solutions we have already discussed.

20.6.2 Computational Equations

20.6.2a Definition of Notation

$\underline{a}$ is an arbitrary factor loading matrix.

$\underline{Z}$ is the normalized data matrix.

$\underline{Y}$ is the arbitrary factor score matrix.

20.6.2b The Equations

$$G = a' a \tag{20.6.1}$$

$$B = a G^{-1} \tag{20.6.2}$$

$$Y = Z B \tag{20.6.3}$$

20.6.3 Computational Instructions. We begin with any arbitrary factor loading matrix.

Eq. (20.6.1) gives the minor product moment of the arbitrary factor loading matrix.

The next step is indicated by Eq. (20.6.2), which is the factor loading matrix postmultiplied by the inverse of the matrix in Eq. (20.6.1). This is the matrix which transforms the data matrix to the factor score matrix.

Eq. (20.6.3) shows the least square factor score matrix as the product of the data matrix postmultiplied by the transformation matrix of Eq. (20.6.2).

20.6.4 Numerical Example. We use the same data and correlation matrices as in the previous sections. We also use the centroid factor matrix of Section 20.3.4.

Table 20.6.1 shows the minor product moment matrix of the centroid factor matrix for three factors.

Table 20.6.2 gives the inverse of the matrix in Table 20.6.1.

Table 20.6.3 gives the product of the natural order of the matrix in Table 20.6.1 postmultiplied by the matrix of Table 20.6.2. This gives the matrix for transforming the data matrix to the least square factor score matrix.

Table 20.6.4 gives the least square factor score matrix.

Table 20.6.5 gives the minor product moment of the least square factor score matrix. This is not an identity matrix, nor should it be so, unless the factor loading matrix consists of basic structure factors.

Table 20.6.1 - Minor Product Moment of Centroid Factor Loading Matrix for Third Factors

3.730	0.170	0.052
0.170	2.017	0.156
0.052	0.156	1.280

Table 20.6.2 - Inverse of Minor Product Moment of Factor Loading Matrix

0.269	-0.022	-0.008
-0.022	0.502	-0.060
-0.008	-0.060	0.789

Table 20.6.3 - Matrix for Transforming Data Matrix to Factor Score Matrix

0.164	0.188	0.327
0.171	0.185	0.313
0.187	0.099	0.333
0.181	-0.303	-0.134
0.161	-0.374	0.112
0.172	-0.353	0.127
0.169	0.151	-0.399
0.160	0.190	-0.411
0.190	0.090	-0.314

Table 20.6.4 - Least Square Factor Score Matrix

0.611	-0.540	0.100
0.302	0.731	0.429
0.319	0.316	-0.812
-0.316	-0.065	-0.208
-0.400	-0.007	-0.109
-0.321	-0.031	-0.083
-0.147	-0.002	0.174
-0.143	0.025	0.299
-0.092	0.018	0.037
0.061	-0.123	0.112
0.124	-0.104	0.012
0.005	-0.215	0.040

Table 20.6.5 - Minor Product Moment of Least Square Factor Score Matrix

0.998	-0.002	0.004
-0.002	1.004	0.004
0.004	0.004	1.050

20.7 The Image Analysis Factor Score Matrix

20.7.1 Characteristics of the Method. When the image analysis approach to factor analysis is used, we may employ any of the factoring methods discussed so far: the principal axis, the group centroid, the centroid, the multiple group, or other methods. Furthermore, we may also use any of the metricizing methods of Chapters 13 and 15. We shall in our description of computational procedures indicate a scaling diagonal. In particular, this may be an identity matrix.

20.7.2 Computational Equations

20.7.2a Definition of Notation

$\underline{R}$ is the correlation matrix.

$\underline{D}$ is an attribute scaling matrix.

$\underline{t}$ is a triangular matrix.

$\underline{Y}$ is the image factor score matrix.

$\underline{Z}$ is the normalized data matrix.

20.7.2b The Equations

$$M = (I - R^{-1} D_{R^{-1}}^{-1}) D \tag{20.7.1}$$

$$G = D (R - 2 D_{R^{-1}}^{-1} + D_{R^{-1}}^{-1} R^{-1} D_{R^{-1}}^{-1}) D \tag{20.7.2}$$

$$C = L' G L \tag{20.7.3}$$

$$t\,t' = C \tag{20.7.4}$$

$$B = M\,(L\,t'^{-1}) \tag{20.7.5}$$

$$Y = Z\,B \tag{20.7.6}$$

20.7.3 Computational Instructions. We shall assume the correlation matrix given. We then calculate a matrix M as indicated in Eq. (20.7.1). This will be recognized as the matrix which transforms the data matrix to the image of the data matrix. The matrix on the extreme right of the right hand side of the equation is a scaling diagonal. It may be chosen according to one of the methods suggested in Chapter 16, or it may be determined according to the self-scaling procedures of Chapter 15.

We then calculate the image covariance scaled matrix, as in Eq. (20.7.2). The part in parentheses on the right hand side of this equation will be recognized as the standard covariance image matrix of Guttman (1953). It is pre- and postmultiplied by the diagonal scaling matrix of Eq. (20.7.1).

Next we calculate the C matrix in Eq. (20.7.3). This is the matrix of Eq. (20.7.2) premultiplied by the transpose of an L matrix and postmultiplied by the natural order of this matrix. This L matrix is of the same order as the factor loading matrix, which presumably has been obtained from the G matrix. In particular, it may be a binary grouping matrix, a matrix of sign vectors for the centroid method, or a principal axis factor loading matrix calculated from the G matrix of Eq. (20.7.2). This depends on which particular type of factor loading matrix one has calculated.

Eq. (20.7.4) shows a triangular factoring of the matrix in Eq. (20.7.3).

Next the $\underline{L}$ matrix is postmultiplied by the inverse of the upper triangular factor of Eq. (20.7.4), and then the matrix of Eq. (20.7.1) is postmultiplied by this product, to give the matrix $\underline{B}$ of Eq. (20.7.5).

The matrix of Eq. (20.7.5) is the transformation matrix which, when applied to the data matrix as in Eq. (20.7.6), yields the factor score matrix. This factor score matrix is orthonormal and of rank reduction form. It is not a least square solution, unless $\underline{L}$ in Eq. (20.7.3) happens to be the basic structure or principal axis factor loading matrix for the covariance matrix $\underline{G}$ in Eq. (20.7.2).

20.7.4 Numerical Example. We use the same data and correlation matrices as in the previous sections, and the grouping matrix of Table 20.4.1. The identity matrix is taken as the scaling diagonal.

Table 20.7.1 gives the image covariance matrix.

Table 20.7.2 gives the image factor score matrix for three factors, as calculated by means of Eqs. (20.7.1) through (20.7.6).

Table 20.7.3 gives the minor product moment of the image factor score matrix. This is an identity matrix to within rounding error, as it should be.

Table 20.7.1 - Image Covariance Matrix

0.733	0.688	0.662	0.120	0.065	0.101	0.294	0.302	0.336
0.688	0.744	0.666	0.147	0.067	0.117	0.316	0.307	0.358
0.662	0.666	0.691	0.215	0.169	0.239	0.288	0.305	0.355
0.120	0.147	0.215	0.511	0.488	0.494	0.218	0.178	0.285
0.065	0.067	0.169	0.488	0.572	0.486	0.161	0.081	0.246
0.101	0.117	0.239	0.494	0.486	0.562	0.160	0.142	0.262
0.294	0.316	0.288	0.218	0.161	0.160	0.481	0.396	0.418
0.302	0.307	0.305	0.178	0.081	0.142	0.396	0.499	0.388
0.336	0.358	0.355	0.285	0.246	0.262	0.418	0.388	0.436

Table 20.7.2 - Image Factor Score Matrix for Three Factors

0.265	0.796	-0.058
0.734	-0.490	0.013
0.018	0.018	0.881
-0.359	-0.112	0.121
-0.383	-0.170	-0.019
-0.316	-0.106	-0.068
-0.025	-0.104	-0.248
0.019	-0.113	-0.331
-0.034	-0.078	-0.051
0.054	0.102	-0.115
0.079	0.111	-0.003
-0.053	0.144	-0.111

Table 20.7.3 - Minor Product Moment of Image Factor Score Matrix

1.000	0.000	-0.004
0.000	1.000	-0.001
-0.004	-0.001	0.999

20.8 Mathematical Proofs

20.8.1 The Centroid Factor Score Matrix

Consider any basic matrix $\underline{L}$ of order $\underline{n} \times \underline{m}$.

Let

$$a_{.i} = \frac{{}_iR\, L_{.i}}{\sqrt{L'_{.i}\, R\, L_{.i}}} \tag{20.8.1}$$

where

$${}_{i+1}R = {}_iR - a_{.i}\, a'_{.i} \tag{20.8.2}$$

Let

$$a = (a_{.1} \dots a_{.m}) \tag{20.8.3}$$

We can prove, by the methods of Chapter 5, from Eqs. (20.8.1) and (20.8.2) that

$$L'\, a = t_a \tag{20.8.4}$$

where $\underline{t_a}$ is lower triangular.

Consider now

$$L'\, R\, L = t_b\, t'_b \tag{20.8.5}$$

and

$$b = {}_1R\, L\, t'^{-1}_b \tag{20.8.6}$$

From Eqs. (20.8.5) and (20.8.6)

$$L' \, b = t_b \tag{20.8.7}$$

From Eqs. (20.8.2) and (20.8.3)

$${}_{m+1}R = R - a \, a' \tag{20.8.8}$$

From Eqs. (20.8.1) through (20.8.4) it can be shown that

$${}_{m+1}R \, L = 0 \tag{20.8.9}$$

From Eqs. (20.8.8) and (20.8.9)

$$L' \, R \, L = L' \, a \, a' \, L \tag{20.8.10}$$

From Eqs. (20.8.5), (20.8.6), and (20.8.7)

$$L' \, R \, L = L' \, b \, b' \, L \tag{20.8.11}$$

From Eqs. (20.8.4), (20.8.7), (20.8.10), and (20.8.11)

$$t_a = t_b \tag{20.8.12}$$

From Eqs. (20.8.4), (20.8.7), and (20.8.12)

$$L' \, a = L' \, b \tag{20.8.13}$$

If $\underline{m} = \underline{n}$ and $\underline{L}$ is nonsingular, we have from Eq. (20.8.13) that $\underline{a} = b$. But from Eqs. (20.8.1) and (20.8.2) the solution for any $a_{\underline{.i}}$ is independent of the solution for any $a_{\underline{.i+k}}$. Also, from Eq. (20.8.6) the solution for any $b_{\underline{.i}}$ is independent of the solution for any $b_{\underline{.i+k}}$. Hence, in general

$$a = b \tag{20.8.14}$$

Consider then the $\underline{n} \times \underline{m}$ matrix $\underline{L}$ of sign vectors for the centroid factor solution such that

$$ {}_{i+1}R = {}_{i}R - a_{.i}\, a'_{.i} \qquad (20.8.15) $$

where

$$ a_{.i} = \frac{{}_{i}R\, L_{.i}}{\sqrt{L'_{.i}\, R\, L_{.i}}} \qquad (20.8.16) $$

From Eqs. (20.8.3) and (20.8.4) we can express the centroid factor matrix as a function directly of the correlation matrix by

$$ a = R\, L\, t'^{-1} \qquad (20.8.17) $$

where

$$ t\, t' = L'\, R\, L \qquad (20.8.18) $$

We may now let $\underline{L}$ and $\underline{L'\, Z'}$ be the right and left arbitrary multipliers in the rank reduction equation and write

$$ E = Z - Z\, L\, [(L'\, Z')\, Z\, L]^{-1}\, (L'\, Z')\, Z \qquad (20.8.19) $$

If

$$ R = Z'\, Z \qquad (20.8.20) $$

we have from Eqs. (20.8.18), (20.8.19), and (20.8.20)

$$ E = Z - (Z\, L\, t'^{-1})\, (t^{-1}\, L'\, R) \qquad (20.8.21) $$

We now consider the general lower rank approximation form to the data matrix

$$E = Z - Y a' \tag{20.8.22}$$

If $\underline{a}$ is given by Eq. (20.8.17), then from Eq. (20.8.21)

$$Y = Z L t'^{-1} \tag{20.8.23}$$

and from Eqs. (20.8.18), (20.8.20), and (20.8.23)

$$Y' Y = I \tag{20.8.24}$$

If now we have an orthonormal transformation

$$b = a H \tag{20.8.25}$$

we consider $\underline{W}$ such that

$$W b' = Y a' \tag{20.8.26}$$

From Eqs. (20.8.25) and (20.8.26)

$$W = Y H' \tag{20.8.27}$$

which is a simple structure factor score matrix. Since $\underline{H}$ is orthonormal by definition, we have from Eqs. (20.8.24) and (20.8.27)

$$W' W = I \tag{20.8.28}$$

If the simple structure transformation is not orthonormal, we use $\underline{h}$ instead of $\underline{H}$ and show that

$$w = Y h^{-1} \tag{20.8.29}$$

and the covariance matrix $\underline{C_w}$ for $\underline{w}$ is

$$C_w = (h' h)^{-1} \tag{20.8.30}$$

or if we let $\underline{r}$ be the correlation of the simple structure reference axes, then

$$C_w = r^{-1} \tag{20.8.31}$$

For computational purposes we have from Eq. (20.8.4)

$$t' = a' L \tag{20.8.32}$$

$$Y = Z (L t'^{-1}) \tag{20.8.33}$$

Then

$$W = Z (L (t'^{-1} H')) \tag{20.8.34}$$

for the orthonormal transformation, and

$$w = Z (L (t'^{-1} h^{-1})) \tag{20.8.35}$$

20.8.2 The Multiple Group Factor Score Matrix

Suppose we let $\underline{f}$ be the binary grouping matrix for the multiple group method of factor analysis. The factor loading matrix given in Chapter 6 is

$$a = R f t'^{-1} \tag{20.8.36}$$

where

$$t\,t' = f'\,R\,f \tag{20.8.37}$$

Then by the methods of the preceding section we have for the rank reduction type multiple group factor score matrix

$$Y = Z\,f\,t'^{-1} \tag{20.8.38}$$

Since

$$R = Z'\,Z \tag{20.8.39}$$

we can readily see from Eqs. (20.8.37), (20.8.38), and (20.8.39) that

$$Y'\,Y = I \tag{20.8.40}$$

We also show that $\underline{Y}$ can be computed directly from the data matrix. We let

$$U = Z\,f \tag{20.8.41}$$

$$t\,t' = U'\,U \tag{20.8.42}$$

From Eqs. (20.8.37), (20.8.38), (20.8.39), (20.8.41), and (20.8.42)

$$Y = U\,t'^{-1} \tag{20.8.43}$$

We may now use Eqs. (20.8.36), (20.8.38), and (20.8.43) to calculate $\underline{a}$ from the factor score matrix $\underline{Y}$, thus

$$a = Z'\,Y \tag{20.8.44}$$

The same procedure could have been used to calculate the centroid factor score matrix directly from the data matrix, having given the matrix L of sign vectors. It will be recalled that in Chapter 5 the successive residual matrices were required from which to calculate the successive sign vectors, and hence they are not available in advance. It is of interest to note that a in Eq. (20.8.44) is precisely a matrix of the correlations of the factor scores with the variables, since both Z and Y are in standard measures. For the case of orthonormal transformations to simple structure we have

$$W = Y H \tag{20.8.45}$$

and for oblique transformations we have

$$w = Y h'^{-1} \tag{20.8.46}$$

For the computational equations we have Eqs. (20.8.41), (20.8.42), and

$$W = U (t'^{-1} H) \tag{20.8.47}$$

and

$$w = U (t'^{-1} h'^{-1}) \tag{20.8.48}$$

for the orthonormal and oblique transformations, respectively.

20.8.3 The Principal Axis Factor Score Matrix

The basic structure or principal axis factor score matrix is well known to be

$$Y = Z Q \Delta^{-1} \tag{20.8.49}$$

or, if the basic structure of $\underline{Z}$ is

$$Z = P \Delta Q' \tag{20.8.50}$$

then

$$Y = P \tag{20.8.51}$$

For the orthonormal and oblique transformation, respectively, we have

$$W = P H \tag{20.8.52}$$

and

$$w = P h'^{-1} \tag{20.8.53}$$

For the covariance matrices of Eqs. (20.8.51), (20.8.52), and (20.8.53) we have

$$C_Y = I \tag{20.8.54}$$

$$C_W = I \tag{20.8.55}$$

$$C_w = (h' h)^{-1} = r^{-1} \tag{20.8.56}$$

20.8.4 The Least Square Factor Score Matrix

Given the arbitrary factor loading matrix $\underline{a}$ and the residual factor score matrix

$$E = Z - Y a' \tag{20.8.57}$$

the solution for $\underline{Y}$ which minimizes $\underline{\mathrm{tr}\ E'E}$ is well known to be

$$Y = Z\, a\, (a'\, a)^{-1} \tag{20.8.58}$$

If

$$a = R\, L\, t'^{-1} \tag{20.8.59}$$

where $\underline{L}$ is arbitrary and

$$R = Z'\, Z \tag{20.8.60}$$

$$t\, t' = L'\, R\, L \tag{20.8.61}$$

then from Eqs. (20.8.58) through (20.8.61)

$$C_Y = Y'\, Y = t'\, (L'\, R^2\, L)^{-1}\, (L'\, R^3\, L)\, (L'\, R^2\, L)^{-1}\, t \tag{20.8.62}$$

If

$$R = Q\, \Delta^2\, Q' \tag{20.8.63}$$

and

$$L = Q\, \Delta \tag{20.8.64}$$

then it can be shown that

$$t = t' = \Delta^2 \tag{20.8.65}$$

and therefore also that

$$C_Y = I \tag{20.8.66}$$

There appears, however, to be no simple expression for C_{Y} for other L matrices, such as the centroid sign matrix, or the binary grouping matrix f.

If we have the orthonormal or oblique transformations H or h, respectively, then

$$W = Y H \tag{20.8.67}$$

and

$$w = Y h'^{-1} \tag{20.8.68}$$

For the case of Eq. (20.8.64) we have

$$C_W = I \tag{20.8.69}$$

$$C_w = r^{-1} \tag{20.8.70}$$

Otherwise,

$$C_W = H' C_Y H \tag{20.8.71}$$

and

$$C_w = h^{-1} C_Y h'^{-1} \tag{20.8.72}$$

as can be seen from Eqs. (20.8.62), (20.8.67), and (20.8.68).

20.8.5 The Image Analysis Factor Score Matrix

Let

$$U = Z (I - R^{-1} D_R^{-1}) D \tag{20.8.73}$$

be the image data matrix scaled with D according to one of the procedures of Chapter 15.

We let

$$M = (I - R^{-1} D_R^{-1}) D \tag{20.8.74}$$

and the scaled image covariance matrix be

$$C_U = U' U \tag{20.8.75}$$

Let L be a matrix of the order of the factor matrix. As a special case, it may be a centroid sign matrix, a binary grouping matrix, or the principal axis factor matrix of C_U. In all cases, we know by the previous methods that the rank reduction type factor score matrix is as follows.

$$u = U L \tag{20.8.76}$$

$$t t' = u' u \tag{20.8.77}$$

$$Y = u t'^{-1} \tag{20.8.78}$$

We let

$$U = P \Delta Q' \tag{20.8.79}$$

and

$$L = Q \Delta \tag{20.8.80}$$

that is, L is the principal axis factor matrix of C_U. Then it can readily be shown from Eqs. (20.8.76) through (20.8.80) that

$$Y = P \tag{20.8.81}$$

$$W = P\ H \tag{20.8.82}$$

$$w = P\ h'^{-1} \tag{20.8.83}$$

and

$$C_Y = I \tag{20.8.84}$$

$$C_W = I \tag{20.8.85}$$

$$C_w = (h'\ h)^{-1} = r^{-1} \tag{20.8.86}$$

For the general case of L we have computationally

$$t\ t' = L'\ D\ (R - 2\ D_R^{-1}{}^{-1} + D_R^{-1}{}^{-1}\ R^{-1}\ D_R^{-1}{}^{-1})\ D\ L \tag{20.8.87}$$

$$B = M\ (L\ t'^{-1}) \tag{20.8.88}$$

$$Y = Z\ B \tag{20.8.89}$$

$$W = Z\ (B\ H) \tag{20.8.90}$$

$$w = Z\ (B\ h'^{-1}) \tag{20.8.91}$$

For the general case we also have Eqs. (20.8.84), (20.8.85), and (20.8.86).